LEAD COMPOUND DISCOVERY FOR MYOTONIC DYSTROPHY

BY

YEN-JUN HO

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Chemistry
in the Graduate College of the
University of Illinois-Urbana-Champaign, 2013

Urbana, Illinois

Advisor

Professor Steven C. Zimmerman

ABSTRACT

Myotonic dystrophy is a debilitating genetic disorder which currently does not have a therapeutic treatment. It is understood that CTG expansions lead to formation of stable poly(CUG) mRNA which mislocalize splicing factors such as MBNL1 and lead to missplicing in the cell. One therapeutic strategy is to target such mutant mRNA with small molecules to prevent the sequestration of splicing factors which will prevent the multiple missplicing events in the cell and reverse the symptoms of the disorder. In the following thesis I describe my work in identifying small molecule inhibitors of the RNA-protein complex through the optimization of gel shift assays for characterization of rationally-designed compounds and the development of a fluorescence anisotropy assay for a high-throughput screening of the NCI Diversity Set III compound library. From such studies I was able to identify several lead compounds that are successful at inhibiting the pathological nuclear aggregation.

ACKNOWLEDGEMENTS

I would like to thank my advisors Professor Steven C. Zimmerman and Professor Anne M. Baranger for giving me guidance and support. Their encouragement and advice inspired me to stay motivated and get through the difficult times in the research project.

I want to thank the Stacie Richardson, Chun-Ho Wong, John Craffey, and Lien Nguyen for their help with the project and the thoughtful and often fruitful scientific discussion which shaped my work. My undergraduate student John Craffey was great and without him the fluorescence anisotropy assay would not have worked. I also want to thank the rest of the graduate students who worked on the myotonic dystrophy project; Yuan, Amin, Long, Jessie, and Tim who assisted with the project. The Zimmerman and Baranger group members have also been great in offering advice and assistance.

Finally, I want to thank my family for supporting me through these years. They were always there for me when I needed them.

TABLE OF CONTENTS

CHAPTER 1

MYOTONIC DYSTROPHY AND THE THERAPEUTIC POTENTIAL OF SMALL MOLECULES

1.1 Overview of Myotonic Dystrophy

Myotonic dystrophy (DM) is a complex genetic disorder with multiple effects across various organ systems.[1] It is the second most common form of muscular dystrophy affecting 1 in 8,000 people worldwide.[2] Features of the disorder include myotonia, muscle wasting, insulin resistance, cardiac conduction defects, cataracts, and cognitive dysfunction.[3] There are two forms of DM based on the location and type of nucleotide expansion. DM1 is the result of CTG repeat expansions in the 3'-UTR of the *DMPK* (dystrophia myotonica protein kinase) gene, whereas DM2 resulted from a tetranucleotide CCTG repeat expansion in intron 1 of the ZNF9 (zinc-finger 9) gene.[4] The genomic instability is hereditary and the length of the repeated region positively correlates with the severity of the disorder.[5] For DM1, repeat lengths greater than 50 results in the development of symptoms of the disorder.[6] Because of such genetic factors, offspring of DM patients experience earlier onset and more severe symptoms.

An RNA gain-of-function model has been proposed to explain the pathogenesis of DM (Figure 1). Upon transcription of expanded poly(CTG) DNA (Fig. 1a), the nascent poly(CUG) RNA folds into a stable secondary structure that aggregates in the nucleus (Fig. 1b); altering cellular levels of DMPK as well as affecting nuclear mRNA transport.[7-9] More importantly, the poly(CUG) aggregate interacts with various splicing factors, resulting in mislocalization and missplicing of multiple proteins. (Fig. 1c)[10-12] Misspliced proteins result in the multisystemic symptoms of DM.[13]

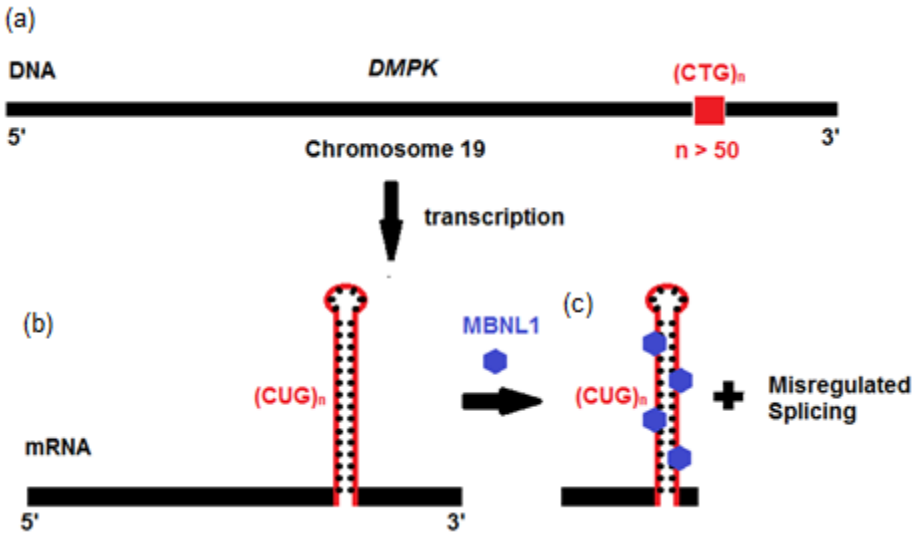


Figure 1. Schematic of RNA gain-of-function model for pathogenesis of myotonic dystrophy

1.2 Trinucleotide Repeat Expansions

The CTG repeats observed in myotonic dystrophy belong to a family of short tandem repeats known as trinucleotide repeats (TNR). Short tandem repeats are quite common in the human genome and constitute 30% of the total DNA.[14,15] Of such sequences, trinucleotide repeats are the most tolerated length of repeats because they do not shift the open-reading frame for protein translation. A survey of the human genome revealed the presence of greater than 32,000 tracts of six or more trinucleotide repeats.[16] Currently there is limited understanding of the function of such repeats but research suggests trinucleotide repeats exist to regulate transcription and translation as expansion leads to various neurological and muscular degenerative disorders.[17] For example, CGG expansion is observed in Fragile X syndrome, CAG expansion is observed in Huntington's disease and 8 other disorders, GAA expansion is observed in Friedreich's Ataxia, GCG expansions is observed in oculopharyngeal muscular dystrophy, and CTG expansion is observed in DM. (Figure 2)[18-20] Although the above mentioned disorders all result from expansion of DNA, the location of the expansion alters disease

pathogenesis. For those expansion in the exon such as Huntington's disease; the pathogenesis is much more direct. Translation of expanded mRNA leads to defective protein with expanded amino acids and leads to loss of protein function. For those expansions that occur in the intron or the untranslated region such as myotonic dystrophy, pathogenesis is often more complex.

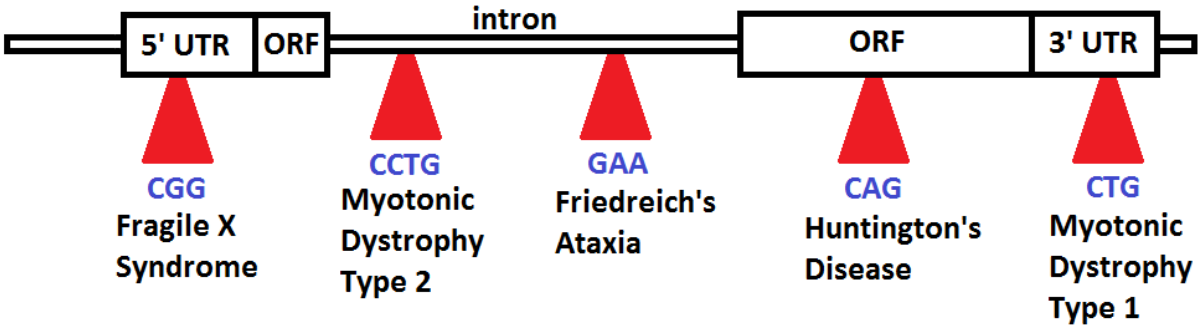


Figure 2. Examples of trinucleotide expansions and the associated genetic disorder

1.3 Mutant DMPK mRNA and MBNL1

As previously mentioned, the poly(CTG) expansions occur in the 3' untranslated region of the *DMPK* gene, thus early studies were focused on understanding the effects of such expansion. In 1995, Singer and coworkers utilized a fluorophore-tagged CAG_{10} antisense oligonucleotide to visualize the location of the poly(CUG) transcript. They reported that the concentration of poly(CUG) transcripts were much higher in DM fibroblasts than normal fibroblasts whereas in the cytoplasm the concentration is the same.[21] . Several years later, a second paper by the group reported that DM myoblasts also had such nuclear foci. They proposed that mutant DMPK mRNA aggregated into nuclear foci to prevent self-export and the export of other mRNAs, leading to disruption of protein synthesis.[22] Reduction in mRNA transport would lead to decreased protein production and potentially DM symptoms. Reddy and coworkers investigated this with mouse models deficient in *DMPK*. They reported that *DMPK*

(-/-) mice developed late-onset, progressive skeletal myopathy that share some pathological features with DM. They suggest DMPK may be necessary for the maintenance of skeletal muscles but loss of DMPK is not the major cause of DM symptoms.[23] This agrees with the observation that the severity of the symptoms correlates with expansion length. If nuclear aggregation leads to DM, then expansion length should not correlate with severity once the threshold expansion size has been reached.

Further studies focused on analyzing the composition of the nuclear aggregates to discover the cause of DM symptoms. A critical observation was made by Thornton and coworkers concerning the presence of various RNA-binding proteins in the nuclear aggregates; including those in the muscle-blind and CUG-BP family of alternative splicing factors.[24] From such observations they hypothesized that mislocalization of alternative splicing factors may be the main cause for DM. In support for such hypothesis, experiments with transgenic mice and fly models both demonstrate that mislocalization of RNA-binding proteins more readily reproduced the symptoms of DM.[25,26] In the drosophila model, Botas and coworkers found that by expressing a non-coding mRNA containing 480 CUG repeats they were able to recreate DM symptoms in the flies. To identify which RNA-binding protein is the most essential, they transfected the DM flies with both MBNL1 and CUGBP1 plasmid and altered the expression level of each protein. It was observed that increased levels of MBNL1 alleviated the RNA-induced toxicity whereas altered levels of CUGBP1 had lesser effects.[25] Mouse models were also studied to confirm the hypothesis. MBNL1 knockout mice ($MBNL1^{\Delta E3/\Delta E3}$) were observed by Swanson to manifest the majority of myotonic dystrophy symptoms.[27] Overexpression of MBNL1 in DM mouse models was again found to alleviate the disorder.[28] The combination of

these results suggested that MBNL1 mislocalization is the main cause for DM and that restoration of native MBNL1 leads to rescue of disease symptoms.

Realizing that MBNL1 is an important protein in the pathogenesis of myotonic dystrophy, multiple research groups focused on understanding its native target and protein structure. Berglund and coworkers published a study in 2007 detailing a comparison between the native substrate of MBNL1 and the secondary structure formed by poly(CUG) repeats.[29] Electrophoretic mobility shift assay (EMSA) was performed to determine the K_d of MBNL1 binding with various lengths of CUG and CCUG repeat constructs, pyrimidine mismatches, and native intronic targets such as cardiac troponin T. Analysis of the results revealed that MBNL1 preferentially binds pyrimidine mismatches, U-U and C-C, with a nanomolar K_d similar to that of the intronic target. Circular dichroism studies and thermal melting data suggests MBNL1 binding requires the bulged stem-loop structure.[30] MBNL1 did not bind true double stranded RNA such as that formed by CUG and CAG.[30] Using a stem-loop RNA construct with a fluorophore at the 5' and a quencher at the 3', Baranger and coworkers observed that upon addition of MBNL1 fluorescence quenching was eliminated. The increase in fluorescence signal suggests that the stem-loop broke apart as a result of MBNL1 binding.[31] A secondary experiment to confirm the observation was done by measuring the binding affinity of MBNL1 to stem-loop RNA with increasing amounts of stem stability. MBNL1 had highest affinity for RNA with weak stem stability and lowest affinity for RNA with strong stability. These findings confirm the importance of bulged stem-loop RNA for MBNL1 binding. A crystal structure of poly(CUG) RNA revealed that it is similar to A-form duplex RNA with a right-handed helical pattern containing bulges at U-U mismatches.[32]

1.4 Potential Therapeutic Approaches

From the RNA gain-of-function model, poly(CTG) DNA is transcribed into mutant DMPK mRNA which forms nuclear foci that mislocalize MBNL1 leading to the symptoms of DM. There are three levels where therapeutic intervention have been suggested to treat the disorder; the DNA level, the RNA level, and the protein level.[33] At the DNA level, one would need to prevent the CTG expansion as well as to contract already expanded CTG tracts. Progress has been limited by the availability of specific DNA binders capable of inducing DNA contraction in non-dividing cells. In a proof-of-concept experiment, Sinden and coworkers have shown that treating DM1 lymphoblast cells with chemotherapeutic agents such as mitomycin C, mitoxantrone, and doxorubicin can lead to CTG repeat contraction.[34] However, such compounds are known to be nonspecific DNA alkylating and intercalating agents with dose-limiting toxicity. Further development to improve selectivity are required if they are to be used in therapy. Antisense oligonucleotides have also been used to stabilize the expansion, but contraction of the expanded DNA with antisense oligonucleotides was not possible.[35] At the protein level, it has been suggested that one can restore the concentrations of RNA-binding proteins. However, that is a difficult task given the protein expression must be tissue selective and carefully controlled. Thus, most research groups have chosen to bypass the DNA and protein level and focus on the RNA level, working to disrupt the formation of the mRNA secondary structure and or preventing MBNL mislocalization.[36,37]

Multiple approaches to target the poly(CUG) repeats have been attempted, including the use of peptides[38], antisense oligonucleotides, and small molecules. With a morpholino poly$(CAG)_{25}$ antisense oligonucleotide Wheeler and coworkers were able to displace mislocalized MBNL1.[39] The oligonucleotide was designed to directly target the CUG repeats

and upon intravenous injection into DM mouse, it rescued the splicing abnormalities associated with the disorder. A more recent example is published by Thornton and coworkers who targeted[40] the intron of the mutant mRNA instead of the poly(CUG) tract. The mechanism of action is believed to result from mRNA silencing through the RNase H-dependent mechanism. The silencing pathway increases the efficiency of the antisense oligonucleotide compared to direct binding to mutant mRNA; less is required to obtain the effect of degrading mutant mRNA nuclear foci. Although these examples demonstrate promise, both peptides and oligonucleotides suffer from the problem of enzymatic degradation, decreased membrane penetration, and poor nuclear retention. Often intravenous injection is required for compound uptake. Because of such limitations, small molecules may be better candidates to target the poly(CUG) RNA due to their smaller size for drug delivery and more established pharmacological properties.

1.5 Published small molecule inhibitors of poly(CUG)-MBNL1 complex

The multiple U-U mismatches in poly(CUG) stem-loop provide binding sites for small molecules due to their rigidity, presence of multiple hydrogen donors and acceptors, and their repeated nature.[41] One of the first small molecule inhibitor published is pentamidine by Berglund and coworkers. (Figure 3)[42] Pentamidine was identified from a screen of twenty-six known nucleic acid binders using gel shift assay. The molecule competed with MBNL1 and disrupted the poly(CUG)-MBNL1 complex with a IC_{50} of 58 μM. In splicing assays, pentamidine was able to partially rescue the missplicing for insulin receptor and cardiac troponin T mRNA. Though promising, pentamidine is dose-limited by high cellular toxicity. There are also concerns that correction of missplicing may result from off-target effects because pentamidine is known to be a non-selective nucleic acid binder.

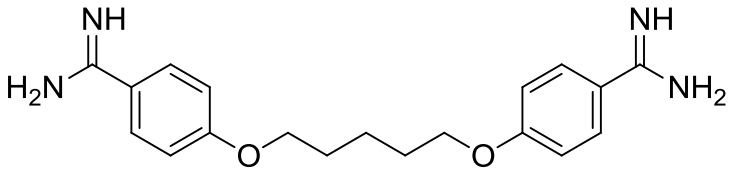


Figure 3. Pentamidine by Berglund and coworkers

To improve selectivity and binding affinity, Disney and coworkers designed a oligomeric molecule based on monomers of a known nucleic acid dye, Hoechst 33258. (Figure 4)[43,44] Each monomer subunit has only weak affinity to the U-U mismatched RNA, but upon optimization of the linker length and conjugating multiple subunits, a pentameric molecule allowed Disney and coworkers to achieve strong binding; IC_{50} of 28 nM. The polymer also improved target selectivity by binding to RNA with consecutive U-U bulges. Such results demonstrate the potential of increasing target specificity and affinity through combination of modular components. However, the drawback of the polymer approach is the large increase in molecular weight of the compound which affects cellular and nuclear uptake.

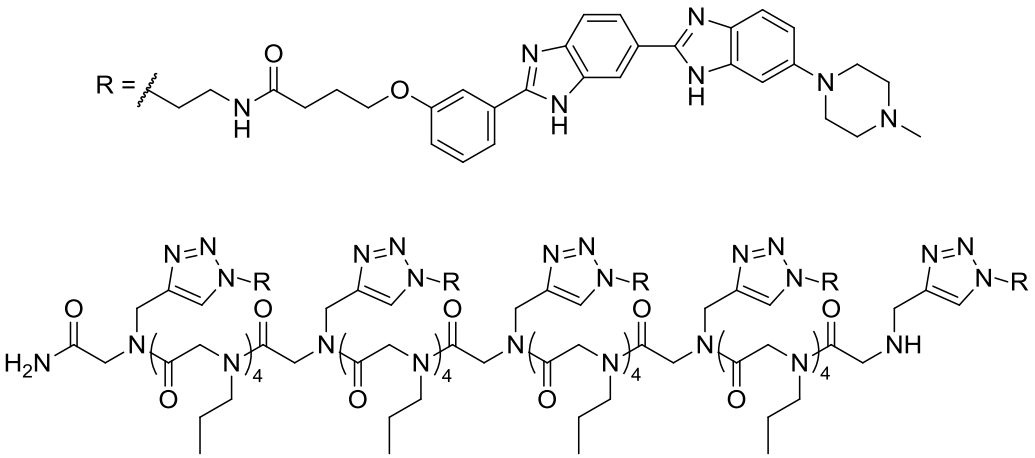


Figure 4. Pentamer of Hoechst33258 (R) by Disney and coworkers

Our own group focused on targeting the RNA with molecules that can hydrogen bond to the U-U bulge. In 2009, we reported on a triaminotriazine-acridine conjugate capable of displacing MBNL1 with an IC_{50} of 46 μM.[45] The design resulted from careful study of the poly(CUG) crystal structure published by Mooers and coworkers[32] which detailed the hydrogen bonding pattern of the U-U mismatch. We reasoned that triaminotriazine can bind between the uracil bases with each face of the molecule fully complementing the hydrogen bonding patterns of the uracil, providing target selectivity. (Figure 5) The conjugated acridine intercalates between RNA bases to afford increased binding affinity as well as stabilize the stem-loop structure which is important in prevent MBNL1 binding. A combination of gel shift assays, isothermal calorimetry measurements, and computer modeling was performed to confirm the proposed binding mode.[46]

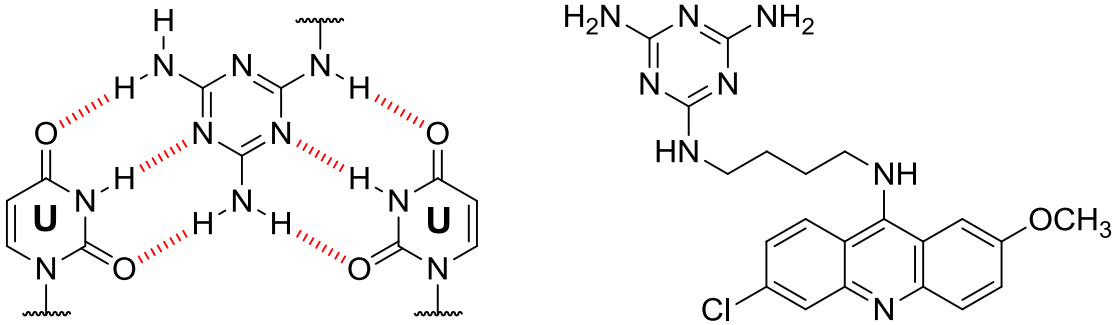


Figure 5. Compounds by Zimmerman and coworkers

1.6 Summary

Myotonic dystrophy is a debilitating genetic disorder with no treatment options. Studies on the pathogenesis of the disorder have found that an RNA gain-of-function is the cause for the disorder. The stable secondary structure formed by the poly(CUG) RNA mislocalizes splicing regulator proteins in the nucleus. Multiple studies suggest MBNL1 is the key splicing regulator

involved in myotonic dystrophy and overexpression of MBNL1 in animal models have reversed the symptoms of DM. As a result, multiple methods were developed to prevent mislocalization of MBNL1. Such methods include the use of synthetic peptides, antisense oligonucleotides, and small molecule inhibitors. Although these methods have shown promise in reversing the symptoms of DM, additional research is necessary to develop them into potential therapeutics.

1.7 References

1. Schoser, B; Timchenko, L. Myotonic dystrophies 1 and 2: complex diseases with complex mechanisms, *Curr Genomics*. **2010**, *11*, 77–90.

2. de Leon MB; Cisneros, B. Myotonic dystrophy 1 in the nervous system: from the clinic to molecular mechanism, *J. Neurosci. Res.* **2008**, *86*, 18–26.

3. Turner, C; Hilton-Jones, D. The myotonic dystrophies: diagnosis and management, *J. Neurol. Neurosurg. Psychiatry* **2010,** *81*, 358–367.

4. Brook, JD; McCurrach, ME; Harley, HG; Buckler, AJ; Church, D; Aburatani, H; Hunter, K; Stanton, VP; Thirion, J-P; Hudson, T; Sohn, R; Zemelman, B; Snell, RG; Rundle, SA; Crow, S; Davies, J; Shelbourne, P; Buxton, J; Jones, C; Juvonen, V; Johnson, K; Harper, PS; Shaw, DJ; Housman, DE. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member, *Cell* **1992**, *68*, 799–808

5. Morales, F; Couto, JM; Higham, CF; Hogg, G; Cuenca, P; Braida, C; Wilson, RH; Adam, B; del Valle, G; Brian, R; Sittenfeld, M; Ashizawa, T; Wilcox, A; Wilcox, DE; Monckton, DG. Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a

heritable quantitative trait and modifier of disease severity, *Human Molecular Genetics* **2012**, *ASAP* published online May 16, 2012.

6. Lee, JE; Cooper, TA. Pathogenic mechanisms of myotonic dystrophy, *Biochem. Soc. Trans.* **2009**, *37*, 1281–1286.

7. Kaliman, P; Llagostera, E. Myotonic dystrophy protein kinase (DMPK) and its role in the pathogenesis of myotonic dystrophy 1, *Cell. Signal.* **2008**, *20*, 1935–1941.

8. Llamusi, B; Artero, R. Molecular effects of the CTG repeats in mutant dystrophia myotonica protein kinase gene, *Curr. Genomics.* **2008**, *9*, 509–516.

9. Cho, DH; Tapscott, SJ. Myotonic dystrophy: emerging mechanisms for DM1 and DM2, *Biochim. et Biophys. Acta* **2007**, *1772*, 195–204.

10. Teplova, M; Patel, DJ. Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1, *Nat. Struct. Molec. Biol.* **2008**, *18*,1364–1377.

11. Yuan, Y; Compton, SA; Sobczak, K; Stenberg, MG; Thornton, CA; Griffith, JD; Swanson, MS. Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs, *Nucl. Acids Res.* **2007**, *35*, 5474–5486.

12. Timchenko, LT; Miller, JW; Timchenko, DD; Datar, KV; Lin, L; Roberts, R; Caskey, CT; Swanson, MS. Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy, *Nucleic. Acids Res.* **1996**, *24*, 4407–4414.

13. Klein, AF; Gasnier, E; Furling, D. Gain of RNA function in pathological cases; focus on myotonic dystrophy, *Biochimie* **2011**, *11*, 2006–2012.

14. Tautz, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers, *Nucleic. Acids Res.* **1989**, *17*, 6463–6471.

15. Toth, G; Gaspari, Z; Jurka, J. Microsatellites in different eukaryotic genomes; survey and analysis, *Genome Res* **2000**, *10*, 967–981.

16. Sobczak, K; Michlewski, G; de Mezer, M; Kierzek, E; Krol, J; Olejniczak, M; Kierzek, R; Krzyzosiak, WJ. Structural diversity of triplet repeat RNAs, *J. Biol. Chem.* **2010**, *285*, 12755–12764.

17. Mirkin, SM. Expandable DNA repeats and human disease, *Nature* **2007**, *447*, 932–940.

18. Tassone, F; Iwahashi, C; Hagerman, PJ. FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS), *RNA Biol.* **2004**, *1*, 103–105.

19. Kumari, D; Biacsi, RE; Usdin, K. Repeat expansion affects both transcription initiation and elongation in friedreich ataxia cells, *J. Biol. Chem.* **2011**, *286*, 4209–4215.

20. Punga, T; Buhler, M. Long intronic GAA repeats causing Friedreich ataxia impede transcription elongation, *EMBO Mol. Med.* **2010**, *2*, 120–129.

21. Taneja, KL; McCurrach, M; Schalling, M; Housman, D; Singer, RH. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues, *J. Cell Biol*. **1995**, *128*, 995–1002.

22. Davis, BM; McCurrach, ME; Taneja, KL; Singer, RH; Housman, DE. Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 7388–7393.

23. Reddy, S; Smith, DBJ; Rich, MM; Leferovich, JM; Reilly, P; Davis, BM; Tran, K; Rayburn, H; Bronson, R; Cros, D; Balice-Gordon, RJ; Housman, D. Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy, *Nat. Genet.* **1996**, *13*, 325–334.

24. Mankodi, A; Teng-Umnuay, P; Krym, M; Henderson, D; Swanson, M; Thornton, CA, Ribonuclear inclusions in skeletal muscle in myotonic dystrophy types 1 and 2, *Ann Neurol*,**2003**, *54*, 760–768.

25. de Haro, M; Al-Ramahi, I; De Gouyon, B; Ukani, L; Rosa, A; Faustino, NA; Ashizawa, T; Cooper, TA; Botas, J. MBNL1 and CUGBP1 modify expanded CUG-induced toxicity in a Dropsophila model of myotonic dystrophy type 1, *Hum. Mol. Genet.* **2006**, *15*, 2138–2145.

26. Mankodi, A; Logigian, E; Callahan, L; McClain, C; White, R; Henderson, D; Krym, M; Thornton, CA. Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat, *Science*, **2000**, *289*, 1769–1773.

27. Kanadia, RN; Johnstone, KA; Mankodi, A; Lungu, C; Thornton, CA; Esson, D; Timmers, AM; Hauswirth, WW; Swanson, MS. A muscleblind knockout model for myotonic dystrophy, *Science* **2003**, *302*, 1978–1980.

28. Kanadia, RN; Shin, J; Yuan, Y; Beattie, SG; Wheeler, TM; Thornton, CA; Swanson, MS. Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 11748–11753.

29. Warf, MB; Berglund, JA. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T, *RNA* **2007**, *13*, 2238-2251.

30. Kino, Y; Mori, D; Oma, Y; Takeshita, Y; Sasagawa, N; Ishiura, S. Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats, *Hum. Mol. Genet.* **2004**, *13*, 495–507.

31. Fu, Y; Ramisetty, SR; Hussain, N; Baranger, AM. MBNL1-RNA recognition; contributions of MBNL1 sequence and RNA conformation, *ChemBioChem*, **2012**, *33*, 112–119.

32. Mooers, BHM; Logue, JS; Berglund, JA, The structural basis of myotonic dystrophy from the crystal structure of CUG repeats, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16626–16631.

33. Mulders, SAM; van Engelen, BGM; Wieringa, B; Wansink, G. Molecular therapy in myotonic dystrophy; focus on RNA gain-of-function, *Hum. Mol. Genet.* **2010**, *19*, 90–97.

34. Hashem, VI; Pytlos, MJ; Klysik, EA; Tsuji, K; Khajav, M; Ashizawa, T; Sinden, RR. Chemotherapeutic deletion of CTG repeats in lymphoblast cells from DM1 patients, *Nucleic. Acids Res.* **2004**, *32*, 6334–6346.

35. Nakamori, M; Gourdon, G; Thornton, CA. Stabilization of expanded (CTG) •(CAG) repeats by antisense oligonucleotides, *Mol. Ther.* **2011**, published online doi:10.1038/mt.2011.191

36. Le Roy, F; Charton, K; Lorson, CL; Richard, I. RNA-targeting approaches for neuromuscular diseases, *Trends Mol. Med.* **2009**, *15*, 580–591.

37. Muntoni, F; Wood, MJA. Targeting RNA to treat neuromuscular disease, *Nat Rev Drug Discov.* **2011**, *10*, 621–637.

38. Artero, RD; Garcia-Lopez, A; Llamusi, B; Orzaez, M; Perez-Paya, E. In vivo discovery of a peptide that prevents CUG-RNA hairpin formation and reverses RNA toxicity in myotonic dystrophy models, *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 11866–11871.

39. Wheeler, TM; Sobczak, K; Lueck, JD; Osborne, RJ; Lin, X; Dirksen, RT; Thornton, CA. Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA, *Science* **2009**, *325*, 336–339.

40. Wheeler, TM; Leger, AJ; Pandey, SK; MacLeod, AR; Nakamori, M; Cheng, SH; Wentworth, BM; Bennett, CF; Thornton, CA. Targeting nuclear RNA for in vivo correction of myotonic dystrophy, *Nature*, **2012**, *488*, 111–117.

41. Kumar, A; Park HJ; Fang, P; Parkesh, R; Guo, M; Nettles, KW; Disney, MD. Myotonic dystrophy type 1 RNA crystal structures reveal heterogeneous 1x1 nucleotide UU internal loop conformations, *Biochemistry* **2011**, *50*, 9928–9935.

42. Warf, B; Nakamori, M; Matthys, CM; Thornton, CA; Berglund, JA. Pentamidine reverses the splicing defects associated with myotonic dystrophy, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 18551–18556.

43. Lee, MM; Childs-Disney, JL; Pushechnikov, A; French, JM; Sobczak, K; Thornton, CA; Disney, MD. Controlling the specificity of modularly assembled small molecules for RNA via ligand module spacing: targeting the RNAs that cause myotonic muscular dystrophy, *J. Am. Chem. Soc.* **2009**, *131*, 17464–17472.

44. Pushechnikov, A; Lee, MM; Childs-Disney, JL; Sobczak, K; French, JM; Thornton, CA; Disney, MD. Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease: application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3, *J. Am. Chem. Soc.* **2009**, *131*, 9767–9779.

45. Zimmerman, SC; Arambula, JF; Ramisetty, SR; Baranger, AM. A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 16068–16073.

46. Zimmerman, SC; Wong, C-H; Richardson, SL; Ho, Y-J; Lucas, AMH; Tuccinardi, T; Baranger, AM. Investigating the binding mode of an inhibitor of the MBNL1-RNA complex in myotonic dystrophy type 1 (DM1) leads to unexpected discovery of a DNA-selective binder, *ChemBioChem*, **2012**, *13*, 2505–2509.

CHAPTER 2

Screening and Identification of Lead Compounds from Rational Design and In Silico Screening

2.1 Introduction

As discussed in the previous chapter, a potential treatment for myotonic dystrophy would be the displacement of mislocalized MBNL from the toxic poly(CUG) repeats with small molecules. To be therapeutically relevant, such small molecules must be selective for the poly(CUG) RNA, bind the RNA with high affinity, and be cell and nucleus permeable. Research in the group has taken two approaches to discover and identify such molecules; rational design by targeting the U-U mismatch with the triaminotriazine motif and an *in silico* pharmacophore-based screen of 170,000 compounds from several small molecule libraries. In the following chapter I detail my work in 1) optimization of electrophoretic mobility shift assay (EMSA) as a method to identify RNA binders, 2) identification of lead compounds from rational design approach, and 3) selection of hit molecules from *in silico* screens.

2.2 Results and Discussion

Optimization of EMSA for Identification of Small Molecule RNA Binders

The electrophoretic mobility shift assay, also known as EMSA or gel shift assay, was initially developed in 1981 by Garner and Revzin for studying the binding between DNA and protein components of the lactose operon system.[1] Swanson and coworkers employed EMSA to analyze the interactions between MBNL1 and both its target pre-mRNA and also toxic poly(CUG) repeats.[2] It is important to note that Swanson utilized a truncated MBNL1N which binds poly(CUG) constructs with binding constants similar to full length MBNL1 yet is much

easier to express using E. coli competent cells.[2] This isoform of MBNL1 is used in our assay along with designed CUG_4 and CUG_{12} stem-loop constructs. (Figure 6) The RNA constructs are radiolabeled with ^{32}P and the proportion of bound and free RNA can be visualized according to their rate of migration in the polyacrylamide gel. The proportion of free RNA allows the quantification of IC_{50} and K_d.

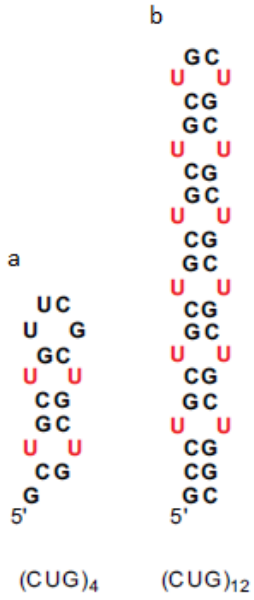


Figure 6. Structure of poly(CUG) construct (a) CUG_4 (b) CUG_{12}

The presence of a super-shifted aggregation band above the protein-RNA complex complicated interpretation of gel shift data, thus initial work focused on its removal. Because the band was first observed for sample lanes with concentrated protein, it was believed to result from MBNL1N self-aggregation. To reduce self-aggregation, the pH of the assay running buffer was decreased from 8.6 to 8.2, which is below the estimated isoelectric point of MBNL1N at 8.3. A positive charge on the protein was believed to reduce protein self-aggregation by initiating binding to negatively charged RNA. CUG_4 and CUG_{12} annealing procedures were also changed

with the goal of favoring the formation of stem-loop RNA over duplex RNA. Fast folding procedures of heat shocking at 90 °C for five minutes followed by five minutes on ice were changed to slow folding procedures where the RNA is placed in 90 °C water and slowly cooled over 90 minutes to room temperature. The combination of these changes along with the addition of glycerol in the gel matrix to affect pore size greatly reduced the presence of aggregation bands. (Figure 7)

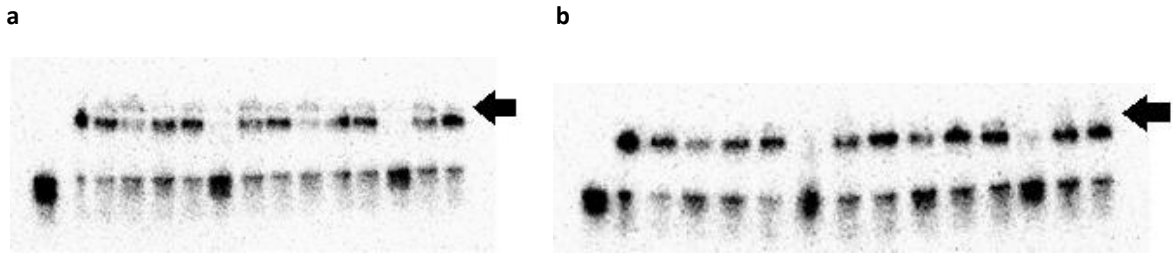


Figure 7. Small molecule inhibition assay to test effect of fast and slow folding with arrow pointing to supershifted band. (a) fast folding (b) slow folding

The aggregation band was reduced but not completely removed until the discovery and removal of a protein contaminant in MBNL1N stock solution. It was observed that MBNL1N expressed and purified at different times demonstrated varying amounts of the aggregation band in the gel shift assay. Also, SDS-PAGE performed with MBNL1N revealed two additional bands upon Coomassie staining. (Figure 8) The two bands are above and below the expected band for MBNL1N with molecular weights of 69 kDa and 26.9 kDa, confirmed by MALDI mass spectrometry. (Figure 9) The contaminant bands were overlooked previously because they were only visible in SDS-PAGE experiments with the use of concentrated protein sample. Incorporating an additional column chromatography step with anionic exchange resin, the contaminants were separated from the protein. Having isolated the contaminants, an equilibrium

binding assay was performed to better understand how they could co-purify with MBNL1N. It was discovered that these contaminants had measureable binding affinity for $(CUG)_{12}$. (Figure 10) The current belief is that the 26.9 kDa band represents a miscleaved GST tag and the 69 kDa contaminant is an MBNL1 isoform because it has a ten-fold weaker binding to CUG_{12}. The lack of aggregation bands in the equilibrium binding assay with the contaminant and MBNL1N suggests that the aggregation originally observed resulted from the mixture of two protein species and was not from assay conditions.

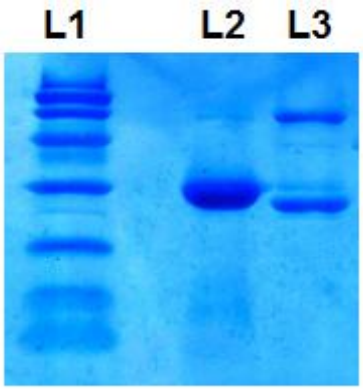


Figure 8. Coomassie stained SDS-PAGE gel.
(L1) protein ladder (L2) purified MBNL1N (L3) contaminants

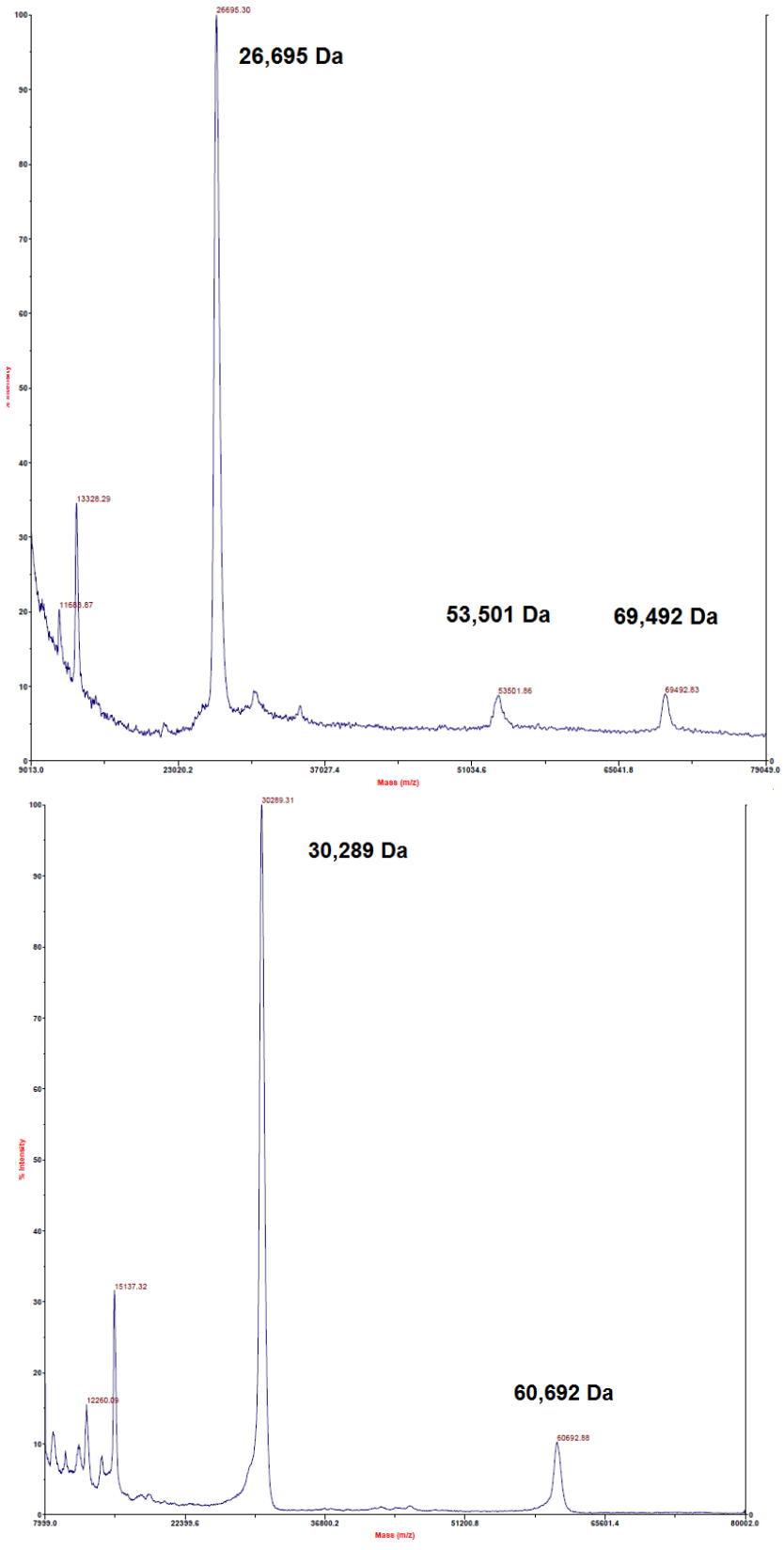


Figure 9. MALDI-MS of isolated impurity (top) and MBNL1N (bottom)

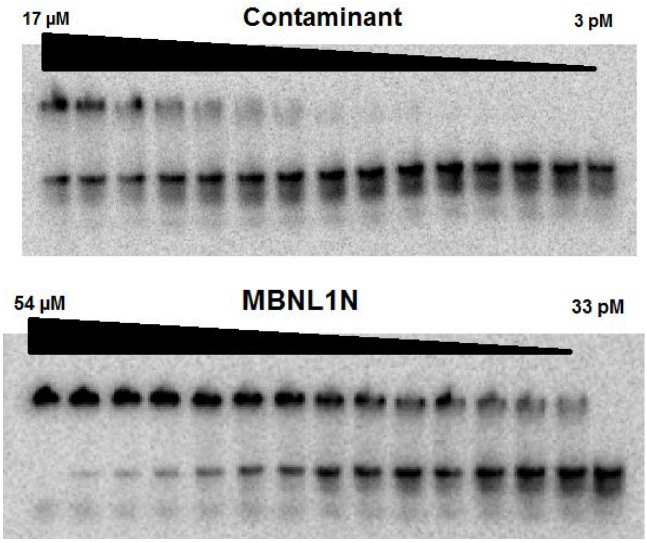
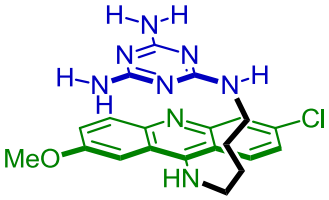


Figure 10. Equilibrium binding assay with (top) Contaminant K_d of 237 nM and (bottom) MBNL1N K_d of 26.6 nM

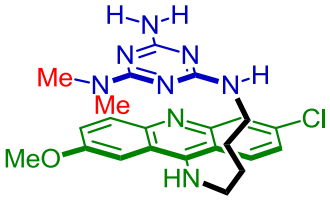
After removing the supershifted aggregation bands, the next focus was on increasing the accuracy of the assay. As explained by Shoichet and coworkers, screening assays are susceptible to identifying false positives resulting from aggregation-based inhibition.[3-5] Some small molecules can exhibit noncompetitive inhibition through the formation of aggregates that prevent the MBNL1 from binding the RNA. The use of detergent molecules such as Triton-X or Tween-20 has been suggested to minimize aggregation-based inhibition. To determine the adequate amount of Triton-X to incorporate, a sample screening assay was performed with known inhibitors and potential inhibitors at different concentrations of Triton-X. The choice of incorporating 0.05% Triton-X in the sample buffer was made based on suggestions from the literature and the observation that this concentration of Triton-X did not alter the inhibitory profiles of the positive control compound. From these studies it was observed that high amounts of Triton-X appear to increase the strength of MBNL1-CUG binding.

Identification of Lead Compounds from Rational Design

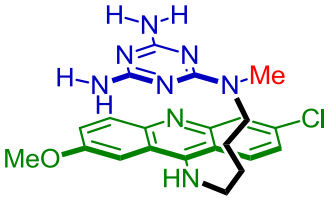
The approach to inhibiting CUG-MBNL binding through rational design of a U-U selective ligands was first reported by Arambula, who used the Janus binding motif of triaminotriazine to target U-U mismatches.[6] Janus wedges are molecules that have two faces available for hydrogen bonding.[7] The combination of triaminotriazine for selectivity and the intercalator acridine for increased nucleic acid binding affinity led to the development of ligand **1**. A series of derivatives has since been synthesized by Wong, a current graduate student in the group, along with second-generation compounds of his own design. (Figure 11)



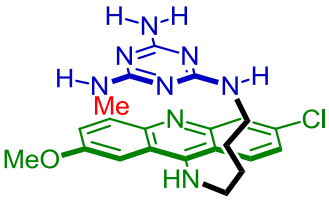
Ligand 1 by Arambula



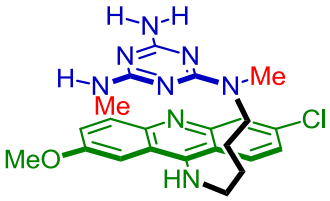
Ligand 2



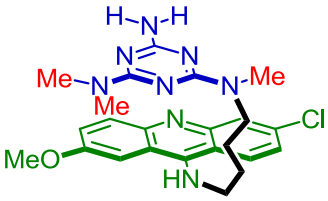
Ligand 3



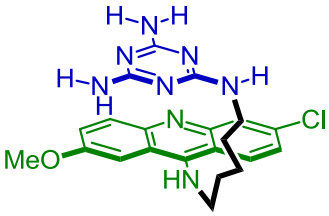
Ligand 4



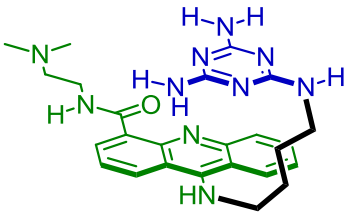
Ligand 5



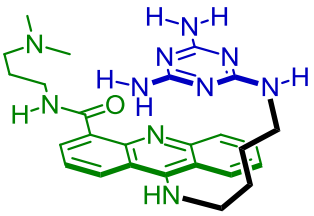
Ligand 6



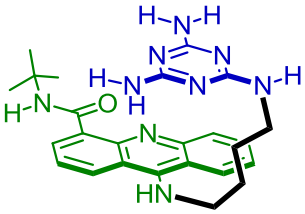
Ligand 7



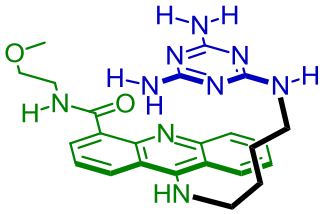
Ligand 8



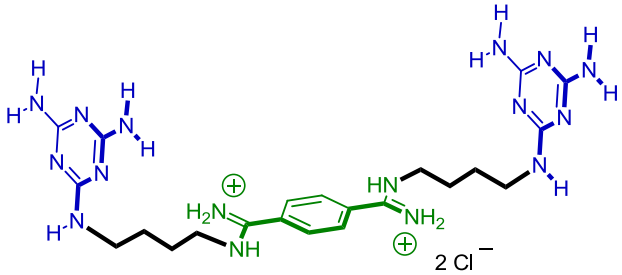
Ligand 9



Ligand 10



Ligand 11



Ligand 12

Figure 11. Summary of triaminotriazine-acridine ligands. All ligands synthesized by Wong unless otherwise noted

After the optimization studies, gel shift assays were used to evaluate the inhibitory effects of rationally-designed compounds. (Figure 12) Ligand **1** is the best characterized compound so it was used as a positive control throughout the majority of EMSA experiments. The other lanes contain derivatives of ligand **1** with methylations in different positions of the melamine for binding mode studies and substituent on the acridine to improve aqueous solubility. A high ratio of MBNL1N to CUG_{12} was utilized to ensure 85% association in the control lane without small molecules. Such stringent condition ensures that only the strongest inhibitors are selected from the assay. From trials performed at 100 μM of small molecule, it was observed that ligand **1** had the best inhibitory activity and the derivatives did not greatly improve the inhibition. Attempts to perform the experiment at higher small molecule concentrations failed due to small molecule-based aggregation in the lanes. Improving aqueous solubility became a major goal for the second generation of triaminotriazine-based ligands. Although inhibition screening experiments did not identify a better derivative, the experimental data gave insight into the binding mode of triaminotriazine-acridine conjugates with U-U mismatches.[8] Ligand **4** contains a single methylation of the amino group and it is the only one of the six methylated derivatives that retains inhibitory activity. This suggests that the proposed Janus wedge binding motif is correct for small molecule binding to RNA as additional methylations on the amine groups prevent the hydrogen bonding and eliminates inhibitory effect.

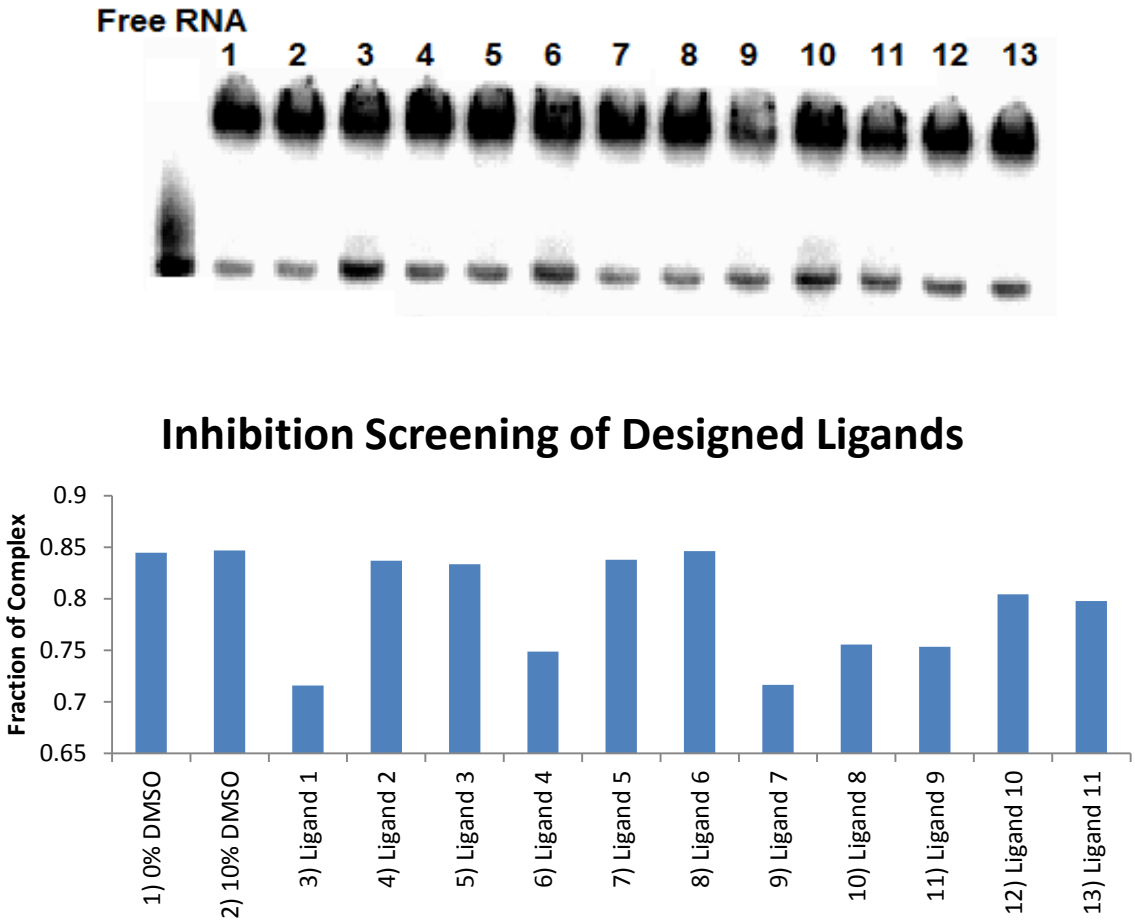


Figure 12. Representative result from a screening experiment.

(Top) Image of screening gel (Bottom) Densitometry of screening experiment

In addition to screening experiments, gel shift assays were utilized to quantify the IC_{50} of the designed compounds. Assuming the designed small molecules behaved as competitive inhibitors, the Cheng-Prusoff equation was used to estimate the K_i for a given compound based on measured IC_{50} and K_d . The K_i is a better measure of a compound's inhibitory effects since it takes into account the strength of the protein-RNA interaction.[9] Because of the high small molecule concentration required for the full displacement of CUG_{12}, reliable IC_{50} and K_i values of ligand **1** derivatives were not obtained. Small molecule-based aggregation occurred in dose-

dependent studies as a result of the compounds' poor aqueous solubility and this prevented it from entering the gel matrix. The second generation compound, ligand **12**, has much better aqueous solubility and its IC_{50} and K_i values have been measured with the gel shift assay. (Figure 13) The low micromolar IC_{50}, 6.7 µM, translates to a K_i of 103 nM which is comparable to the most potent published small molecule inhibitor at 76 ± 15 nM.[10,11] Fluorescence microscopy results visualize the displacement of MBNL from poly(CUG) foci *in vivo*, which provides support for the validity of the gel shift assay as a method to determine inhibitory activity.

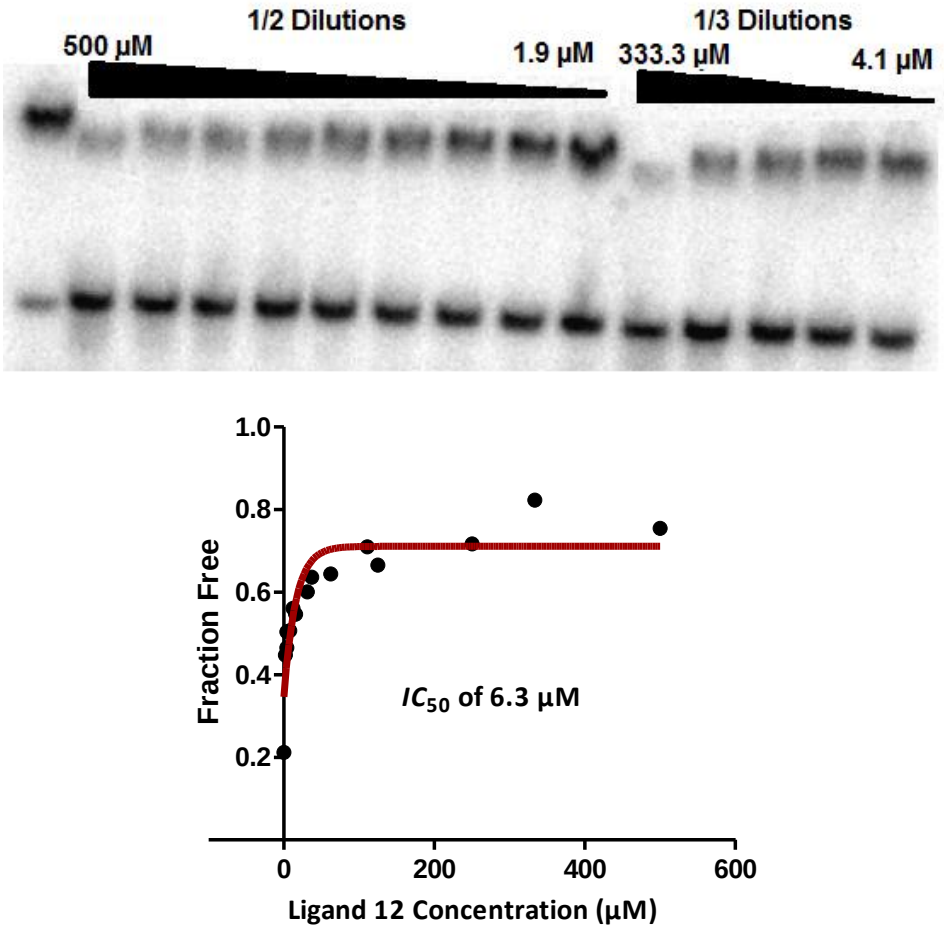


Figure 13. Representative result from dose-dependent experiment
(Top) Representative gel image (Bottom) Densitometry data

Identification of Lead Compounds from *in silico* Pharmacophore Screening

A screening approach was used to identify small molecules that can target the minor groove of the poly(CUG) RNA. A postdoctoral fellow in the lab, Tiziano Tuccinardi, designed an *in silico* glycerol-based pharmacophore screen of 170,000 small molecules from the Chembridge, Marvel, NCI, and high-throughput screening facility at the University of Illinois compound libraries. The glycerol pharmacophore was chosen because the crystal structure submitted by Krzyzosiak and coworkers (PDB ID: 3GLP) contained a glycerol interacting with the minor groove of poly(CUG)$_4$. (Figure 14)[12] Ultimately the goal of the screen is to identify molecular scaffolds which have affinity and selectivity to the poly(CUG) minor groove such that it can be combined with known targeting motifs.

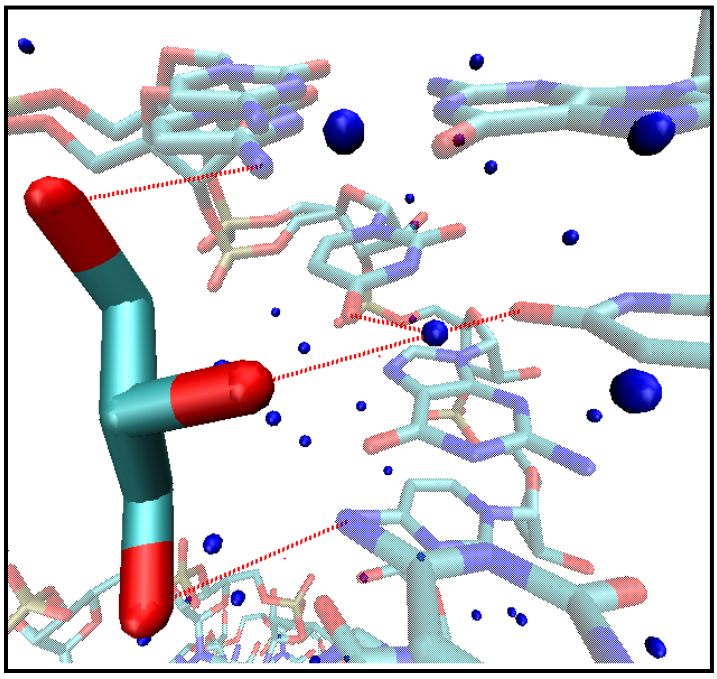
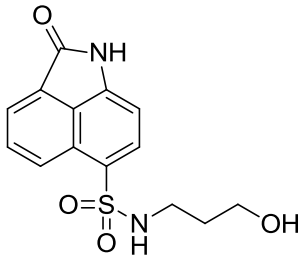
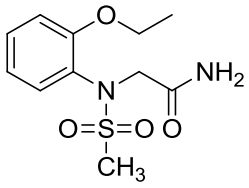


Figure 14. Glycerol bound in minor groove of CUG_4

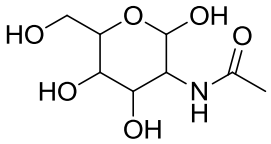
The glycerol-based in silico screen resulted in 117 molecules being identified as potential binders of trinucleotide repeat CUG. Gel shift assays were performed with these identified molecules to determine their inhibitory activity. The preliminary screening experiment utilized a CUG_4 RNA construct to avoid time-consuming assay optimization. From the 117 molecules, 12 hits were identified based on greater than 5% displacement of CUG_4 in comparison to the DMSO control. (Figure 15) Most of the compounds identified only displaced 6-7% of RNA with Marvel 4796 being the most active at 10% displacement. There were concerns about the validity of CUG_4 as a model for poly$(CUG)_n$ because it only has two U-U mismatches upon folding and might not contain enough bases to resemble A-form RNA which is formed by poly$(CUG)_n$. In addition, there were concerns that the UUCG motif in CUG_4 construct may alter MNBL1 binding to CUG.[13] Thus, the screen was repeated with a CUG_{12} RNA construct.



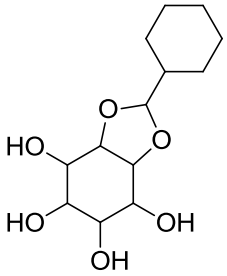
Chembridge 5484017



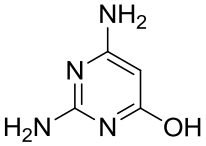
Chembridge 6185525



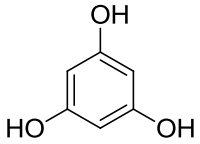
Marvel 1174



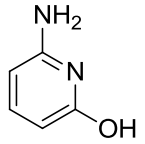
Marvel 1264



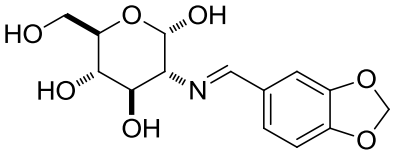
Marvel 4742



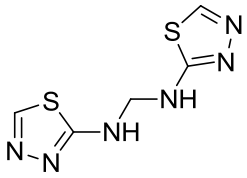
Marvel 4796



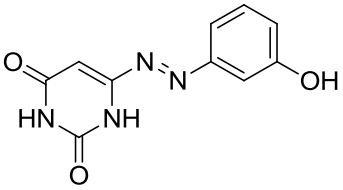
Marvel 531



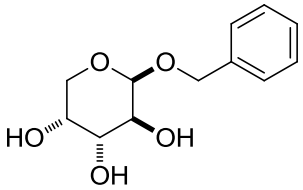
NSC 71426



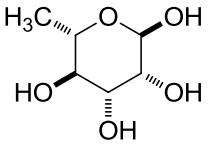
NSC 143019



NSC 194250



NSC 2561

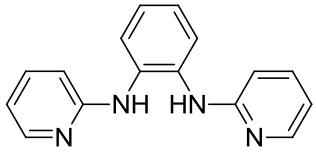


NSC 287050

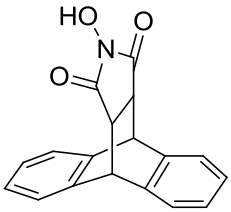
Figure 15. Hit compounds identified with CUG_4 construct

With 6 U-U mismatches upon folding and lacking the UUCG cap motif, CUG_{12} is a better model of toxic RNA than CUG_4. However, assay optimization was much more time-consuming. The screen resulted in identification of 13 lead compounds following the same guidelines as the CUG_4 screen. (Figure 16) Ligand **1**, the positive control displaced an average of 23% relative to the RNA-protein complex while most identified hits averaged 8%. (Figure 17) Surprisingly, none of the compounds were identical from the two screens. This suggested that

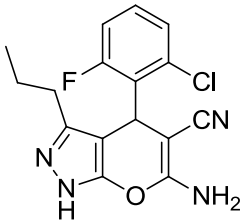
the CUG_4 and CUG_{12} constructs may not form the same helical RNA structure in solution because completely different hits were identified. Because we believed the CUG_{12} construct to better represent poly$(CUG)_n$, we used the 13 hit molecules from the second screen for future analysis.



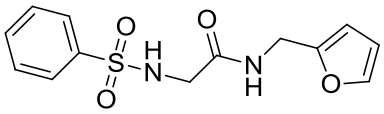
Chembridge 5367911



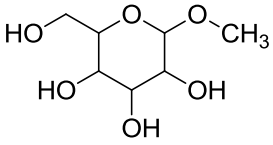
Chembridge 5918715



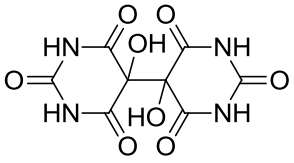
Chembridge 6659423



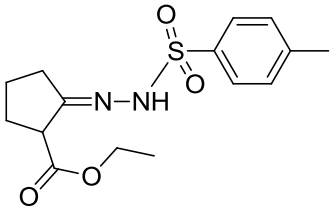
Chembridge 7490819



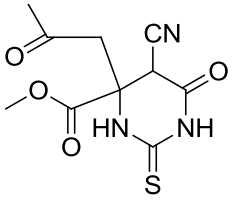
Marvel 4761



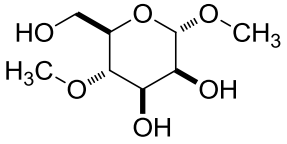
Marvel 4757



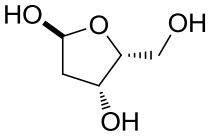
HTSF 903



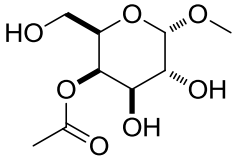
NSC 371679



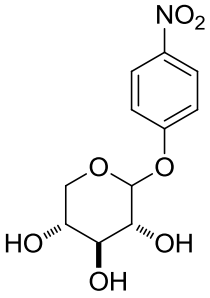
NSC 403459



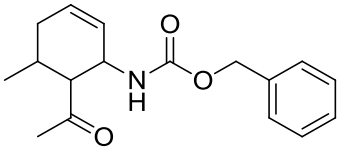
NSC 403510



NSC 403958



NSC 371094



NSC 380247

Figure 16. Hit compounds identified with CUG_{12}

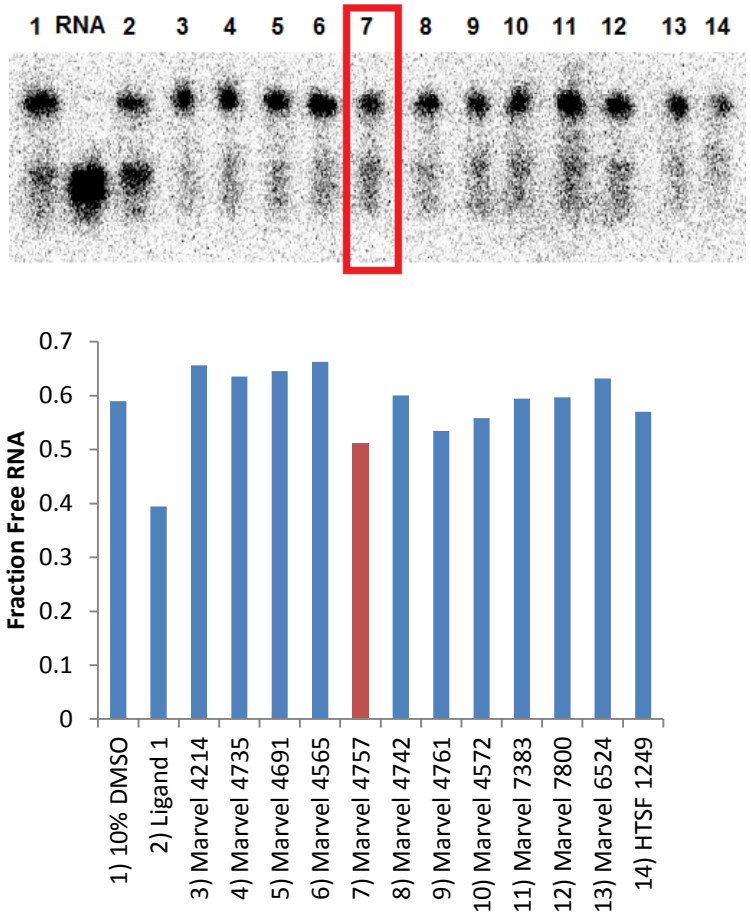


Figure 17. Representative inhibition screening experiment
(Top) gel image with hi compound boxed (Bottom) Densitometry data with hit shown in red

Dose-dependent experiments with the 13 small molecules were inconclusive. We observed that experimental data was highly variable between trials and high small molecule concentrations were required to observe significant inhibition. Eventually it was realized that the sample binding buffer required for gel shift assay affected the results of the study. Glycerol is a density increasing agent for gel shift samples and it is required for the experiment. Because our initial pharmacophore model is based on glycerol binding in the minor groove, additional glycerol in the buffer blocks the intended target site. Experimental protocols describe the use of

1.5 M glycerol per sample and steady-state fluorescence titration measurements with glycerol and TAMRA-CUG_6 performed by a coworker showed a K_d of 1 M for glycerol with the RNA. In other words, greater than half of the potential binding sites on CUG_{12} are occupied by glycerol in the gel shift assay. The assay would be biased towards identifying compounds not sharing the same binding mode as glycerol. There are alternative density-increasing agents one can use but to address this problem a fluorescence-based assay was developed and will be described in chapter 3.

2.3 Conclusions

Through methodical improvements of assay conditions and sample buffer, a gel shift assay was developed into a more precise experiment for determining small molecule inhibition of MBNL1N-poly(CUG) binding. Improvements allowed the direct comparison of ligand **1** derivatives which was used in understanding the binding mode of the compounds as well as obtaining dose-dependent data for second generation compounds. Identification of ligand **12** as a strong inhibitor of RNA-protein interactions validated the rational design approach. Screening experiments used to evaluate the glycerol-based pharmacophore model identified several compounds with observable inhibitor activity. However, due to the necessity of glycerol as a density-increasing agent for the assay, the data obtained are inconclusive. This experience served as motivation for the development of an orthogonal fluorescence-based method for determining small molecule inhibition.

2.4 Materials and Methods

Plasmid

The expression vector pGEX-6p-1/MBNL1N (272 amino acids) was obtained from Maurice S. Swanson.[2] MBNL1N has the four zinc finger motifs of MBNL1 and also a GST-tag at its N-terminal and a His_6-tag at its C-terminal for purification.

MBNL1N Expression and Purification

Bl21 E. coli cells were transformed with the GST-MBNL1N plasmid. A mixture of 50 µl of cell, 1 µL of plasmid, and 0.7 µL of 10% BME was kept on ice for 20 min and then heat shocked for 45–55 sec at 42 °C. After returning to ice for 30 sec, 450 µL of SOC media was added at room temperature. The mixture was incubated in a shaker at 37 °C for 1 h. and then plated on agar and placed in a convection oven at 37 °C overnight for bacterial growth.

Transformed bacteria were induced with IPTG, once the OD_{600} reached 0.6, for 2 h at 37 °C. The solution was centrifuged for 15 min at 10,000 g and the pellet was saved and treated with lysis buffer (tris-Cl pH 8, 10 mM imidazole, 2mg/mL lysozyme, 0.1% Triton X, 0.1% protease inhibitor, 0.5 M NaCl, 2 mM BME, 5%glycerol). Additionally, DNAse 1 was added to the lysate in a ratio of 1 µL of DNAse to 1 mL of lysis buffer. After 6 rounds of sonication at 30 sec each followed by 20 sec ice bath, the sample was centrifuged at 12,000 g for 15 min. The supernatant was filtered with a 0.45 µM Millex Filter (Millipore). One mL of Ni NTA Sepharose (QIAGEN) was added and incubated with lysate for 1 h at 4 °C. The supernatant was removed and the beads are washed with 30 mL of wash buffer 1 (25 mM tris-Cl pH 8, 20 mM imidazole, 0.5 M NaCl, 0.1% Triton X). The protein was eluted with 20 mL of elution buffer

(25 mM tris-Cl pH 8, 250 mM imidazole, 0.5 M NaCl, 0.1% Triton X water). One mL of Glutathione Sepharose 4B (GE Healthcare) was incubated with the eluent for 1 h at 4 °C. The solution was loaded on a column and the beads were washed with 20 mL of wash buffer 2 (25 mM tris-Cl pH 8, 5 mM BME, 300 mM NaCl, 0.1% Triton X). The column was sealed and the beads were suspended in 3 mL of wash buffer 2. To the suspension, 5 mM of fresh BME and 40 units of Precission Protease (GE Healthcare) were added. This solution was kept at 4 °C with constant shaking for 48 h. The suspension was centrifuged at 12,000 g and the supernatant was collected. The solution was loaded into a column with Q Sepharose Fast Flow resin (Sigma-Aldrich) and the protein was eluted with a gradient salt buffer (high salt buffer: 500 mM NaCl, 25 mM tris-Cl pH 7.5, 5 mM BME; low salt buffer: 100 mM NaCl, 25 mM tris-Cl pH 7.5, 5 mM BME). The concentration of protein obtained was determined with a BCA assay (Thermo Scientific) while purity was confirmed with MALDI mass spectrometry and SDS-PAGE.

RNA Purification

$(CUG)_4$ and $(CUG)_{12}$ were purchased from Integrated DNA Technologies. $(CUG)_4$ was not purified while $(CUG)_{12}$ was purified by performing a denaturing polyacrylamide gel electrophoresis with 12% polyacrylamide (National Diagnostics). The gel was pre-run for 30 min at 50 W prior to loading the RNA sample. The RNA sample is mixed 1:1 with formamide buffer (95% fomamide, 10 mM tris-Cl pH 8, 20 mM EDTA, 100 µM bromophenol blue, 100 µM xylene cyanol). Before loading, the sample is heat shocked for 2 min at 90 °C. The experiment is performed at 4 °C with the power at 50 W until the dye has migrated to the bottom quarter of the polyacrylamide gel. Afterwards the gel is removed from the glass plates, observed with UV lamp to mark the RNA region, and that portion is removed with a sterile razor. The gel fragment

is forced through a syringe and 3 mL of 1X TE buffer is added for every 0.5 mL of gel fragment collected. The tube is frozen at -80 °C for 30 min, thawed at 50 °C for 5 min, and soaked at 90 °C for 5 min. It is left on a shaker overnight. The sample is extracted with n-butanol with the upper butanol layer removed after gentle shaking. Two rounds of ethanol precipitation followed. For every 100 µL of sample solution, 30 uL of 3 M NaCl solution was added as well as 3 volumes of cold ethanol. The sample is left in dry ice for 30 min before centrifuged at 13,200 rpm for 30 min at 4 °C. Afterwards ethanol was removed at reduced pressure and G-25 Microspin (GE Healthcare) was used for buffer exchange into water. RNA concentration was determined on Shimadzu UV2450 by measuring the absorbance at 260 nm.

RNA Radiolabeling

$(CUG)_4$ and$(CUG)_{12}$ were labeled with 50 uCi [γ-^{32}P] ATP with T4 Kinase (Invitrogen). The RNA was heat shocked for 5 min at 90 °C prior to the experiment. The reaction mixture, RNA, ATP, kinase and buffer, was placed in a water bath at 37 °C for 2 h and then moved to heat block of 70 °C for 30 min. Chloroform/phenol extraction was performed by diluting the sample to 50 µL with water followed by addition of 50 µL of 25:24:1 phenol/chloroform/isoamyl alcohol. After vortex and centrifuge the lower organic layer is removed. A second extraction with 50 µL of 24:1 chloroform/ethanol is added and after vortex and centrifuge, the bottom organic layer is again removed. Ethanol precipitation is performed twice following chloroform/phenol extraction. The resulting solid is diluted with water and stored at -20° C.

Equilibrium Binding Assay

The RNA was placed in 150 mL of near-boiling water and both left on the bench to return to room temperature for 90 min. Next, the RNA samples were diluted with protein binding buffer (175 mM NaCl, 5 mM $MgCl_2$ 20 mM tris-Cl pH 8, 1.25 mM BME, 12.5% glycerol, 2 mg/mL BSA, 0.1 mg/mL heparin, and 0.05% Triton-X) to 0.4 nM. The protein was serially diluted with protein binding buffer as well. The two were mixed and incubated at room temperature for 20 min. They were then loaded onto pre-chilled 4 °C 6% acrylamide gels in 0.5X tris-borate buffer pH 8.2. The gel was electrophoresised at 180 V for 1 h and then dried and visualized on a Molecular Dynamics Storm Phosphorimager. Apparent K_d was obtained by fitting to the equation fraction RNA bound = $1/(1 + K_d / [\text{Protein}])$ using KaleidaGraph 3.5.

Inhibition Screening Assay

The RNA was prepared in the same fashion as the equilibrium binding assay. The protein concentration used is the 80% bound state relative to the K_d. Loading volumes were 10 μL with 4.5 μL of RNA, 4.5 μL of protein, and 1 μL of small molecule. The small molecule and protein were mixed initially, incubated for 20 min and then the RNA was added followed by 20 min incubation time. The gels were electrophoresised at 180 V for 1 h, dried, and visualized with phosphor imager. Inhibition was determined by densitometry of free to complexed RNA.

IC_{50} Assay

Small molecules were serially diluted from 10 mM and incubated with RNA for 20 min. MBNL1 was added and an incubated for an additional 20 min loading volume of 10 μL was used. The gels were electrophoresised at 180 V for 1 h and then dried and visualized. To determine

IC_{50} values the data was fit to the equation $B = \Delta B \exp((-0.69/IC_{50})C) + B_f$ where B is the volume of the free RNA and ΔB is the difference in the volume of the free RNA band at the lowest and highest concentrations of small molecule, C is the concentration of the small molecule, and IC_{50} is the concentration of small molecule at which 50% of the RNA is dissociated from the protein.

2.5 References

1. Garner, MM; Revzin, A. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system, *Nucleic Acids Res.* **1981**, *9*, 3047–3060.

2. Yuan, Y; Compton, SA; Sobczak, K; Stenberg, MG; Thornton, CA; Griffith, JD; Swanson, MS. Muscleblind-like 1 interacts with RNA hairpins in splicing target and pathogenic RNAs, *Nucl. Acids Res.* **2007**, *35*, 5474–5486.

3. Seidler, J; McGovern, SL; Doman, TN; Shoichet, BK. Identification and prediction of promiscuous aggregating inhibitors among known drugs**,** *J. Med. Chem*. **2003**, *46*, 4477–4486.

4. Feng, BY; Shoichet, BK. A Detergent-Based Assay for the Detection of Promiscuous Inhibitors**,** *Nature Protocols*. **2006**, *1*, 550–553.

5. Feng, BY; Simeonov, A; Jadhav, A; Babaoglu, K; Inglese, J; Shoichet, BK; Austin, CP. **A** High-Throughput Screen for Aggregation-Based Inhibition in a Large Compound Library, *J. Med. Chem*. **2007**, *50*, 2385–2390.

6. Zimmerman, SC; Arambula, JF; Ramisetty, SR; Baranger, AM. A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 16068–16073.

7. Branda, N; Guido, K; Lehn, JM. JANUS WEDGES: a new approach towards nucleobase-pair recognition, *Chem. Comm.* **1996**, 2443-2444.

8. Zimmerman, SC; Wong, C-H; Richardson, SL; Ho, Y-J; Lucas, AMH; Tuccinardi, T; Baranger, AM. Investigating the binding mode of an inhibitor of the MBNL1-RNA complex in myotonic dystrophy type 1 (DM1) leads to unexpected discovery of a DNA-selective binder, *ChemBioChem*, **2012**, *13*, 2505–2509.

9. Cheng, Y-C; Prusoff, WH. Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction, *Biochem. Pharmacol.* **1973**, *23*, 3099 – 3108.

10. Lee, MM; Childs-Disney, JL; Pushechnikov, A; French, JM; Sobczak, K; Thornton, CA; Disney, MD. Controlling the specificity of modularly assembled small molecules for RNA via ligand module spacing: targeting the RNAs that cause myotonic muscular dystrophy, *J. Am. Chem. Soc.* **2009**, *131*, 17464–17472.

11. Pushechnikov, A; Lee, MM; Childs-Disney, JL; Sobczak, K; French, JM; Thornton, CA; Disney, MD. Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease: application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3, *J. Am. Chem. Soc.* **2009**, *131*, 9767–9779.

12. Kiliszek, A; Kierzek, R; Krzyzosiak, WJ; Rypniewski, W. Structural insights into CUG repeats containing the 'stretched U-U wobble': implications for myotonic dystrophy, *Nucl. Acids Res.* **2009**, *37*, 4149–4156.

13. Warf, MB; Berglund, JA. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T, *RNA* **2007**, 13, 2238–2251.

CHAPTER 3

FLUORESCENCE ANISOTROPY FOR IDENTIFICATION OF LEAD COMPOUNDS

3.1 Introduction

In 2006 Shapiro and coworkers described the use of fluorescence anisotropy for small molecule discovery for the vigilin-vitellogenin mRNA complex.[1] The paper demonstrated the feasibility of using anisotropy in a high-throughput format for screening small molecule inhibitors of RNA-protein complexes. However, no such assay has been published for the MBNL1-poly(CUG) system. Compared to EMSA, anisotropy is much less time-consuming and laborious, it is an equilibrium-based measurement, and most importantly can be performed in a high-throughput methodology. Furthermore, a fluorescence anisotropy assay could address the two inherent limitations of the gel shift assay 1) the requirement of a solid gel matrix for sample separation which prevents certain small molecules from being analyzed and 2) the necessity of using glycerol in the sample buffer. This would allow us to complete the identification of lead molecules from the pharmacophore screen and allow for the design of a high-throughput screen.

3.2 Results and Discussion

Experimental Design

Fluorescence anisotropy measures the loss of polarization of the light emitted from the fluorophore, which is affected by the rate of tumbling of the bound complex. Thus, to obtain the best signal difference between bound and free states, one needs the bound state to be much larger than the unbound, which minimizes the tumbling rate. (Figure 18) The RNA construct used for the experiments is TAMRA-CUG_6 (excitation 546 emission 579), which is attached to the 5'-end.

(Figure 19) TAMRA was chosen as the fluorophore for its longer wavelength of excitation and emission that does not overlap with that that of ligand **1**, its derivatives and most small molecules. Instead of using MBNL1N, 30 kDa, the larger GST-MBNL1N, 56 kDa, was selected. The anisotropy buffer utilized was based on protocols from an AlphaScreen assay developed by Thornton and coworkers.[2] However, BSA and Triton-X100 were supplemented to disrupt potential nonspecific small molecule and protein interactions.

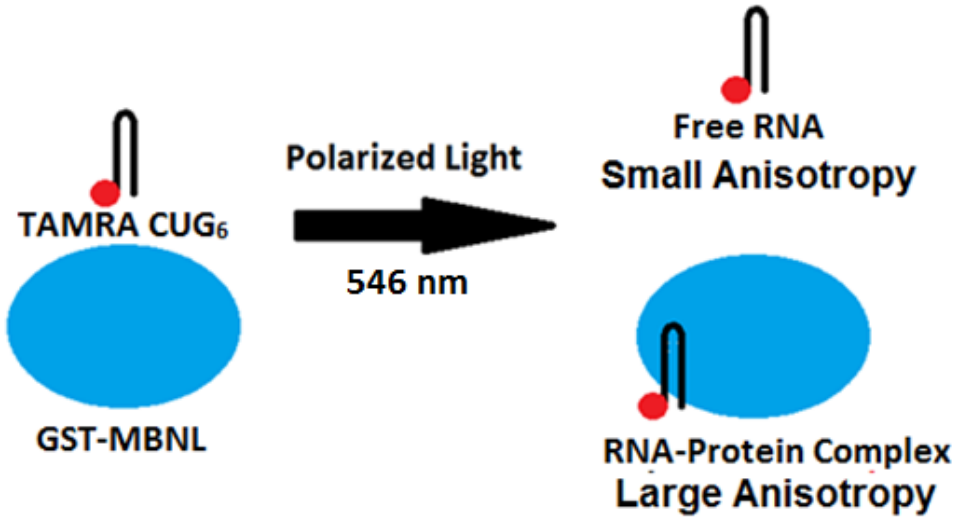


Figure 18. Schematic of fluorescence anisotropy

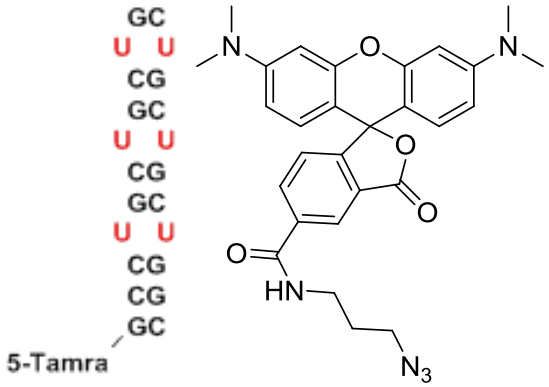


Figure 19. RNA construct (Left) CUG_6 construct (Right) Structure of TAMRA fluorophore

Development of Assay Conditions and Protocols

Although the TAMRA-CUG_6 was easily obtained through commercial sources, GST-MBNL1N was difficult to purify and store. Previously established protocols for expression of MBNL1N, transcribed from the GST-MBNL1N plasmid, utilizes Precision Protease (Amersham) to cleave the GST-tag after affinity chromatography with glutathione resin. Elimination of the protease step to preserve the GST tag reduced the purity of the collected protein. (Figure 20a) MBNL1N and GST is believed to be the major contaminants in the fraction based on SDS PAGE analysis. To increase purity, cell lysate sonication time was reduced and affinity chromotography protocols were altered in comparison to MBNL1N expression to prevent self-cleavage of GST-MBNL1N. Excessive sonication time is known to cause cleavage of GST from fusion proteins.[3] Improved purity of GST-MBNL1N was observed in SDS-PAGE (Figure 20b) and indirectly from stronger binding for GST-MBNL1N and CUG_6 from fluorescence anisotropy experiments.

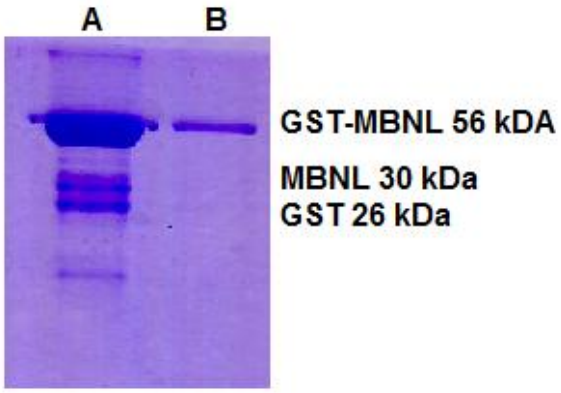


Figure 20. SDS PAGE of collected protein
(a) protein prior to additional purification steps (b) protein post purification

After purification and concentration of GST-MBNL1N fractions, it was observed that protein precipitation had occurred. It is known that GST-tagged proteins form dimers and tetramers and that such association is dependent on concentration. For our purposes we needed the protein sample to be stored at high concentrations to avoid diluting assay buffer conditions. To prevent precipitation, we incorporated protein dialysis into the purification procedure, increased the amount of glycerol in the storage buffer, and maintained the concentration of the protein below 10 μM. Even with these changes we observed protein precipitation over a period of four weeks. Thus, it was decided that GST-MBNL1N be expressed fresh prior to fluorescence anisotropy experiments and screens.

At the completion of these optimization steps we began to test the validity of the assay by performing equilibrium binding assays with GST-MBNL1N and Tamra-CUG_6 as well as inhibition assays with ligand **1** and comparing to published data. Compared to the K_d values obtained with gel shift assays, the K_d values from anisotropy assays were about ten folds higher at 100 nM. It is difficult to determine whether this resulted from different RNA construct or from the difference in assay. For measuring K_d, fluorescence anisotropy data is much more variable than gel shift assays due to the slow degradation of GST-MBNL1N in solution over time.

Ligand **1** was chosen to be the control compound for inhibition studies because its IC_{50} and K_i data is already published by Zimmerman and coworkers.[4] Due to the slowly changing concentration of the GST-MBNL1N stock solution, a K_d was determined prior to each inhibition assay to determine the amount of GST-MBNL1N to use. After three independent experiments we found that based on the anisotropy data the IC_{50} for ligand **1** is 87 ± 25 μM, from which a K_i value of 7 ± 2 μM is calculated. This is similar to the published K_i of 7 ± 1 μM for ligand **1** from

gel shift assays. (Figure 21) This result demonstrated that fluorescence anisotropy is a valid assay to determine small molecule inhibition of poly(CUG)-MBNL1 interactions.

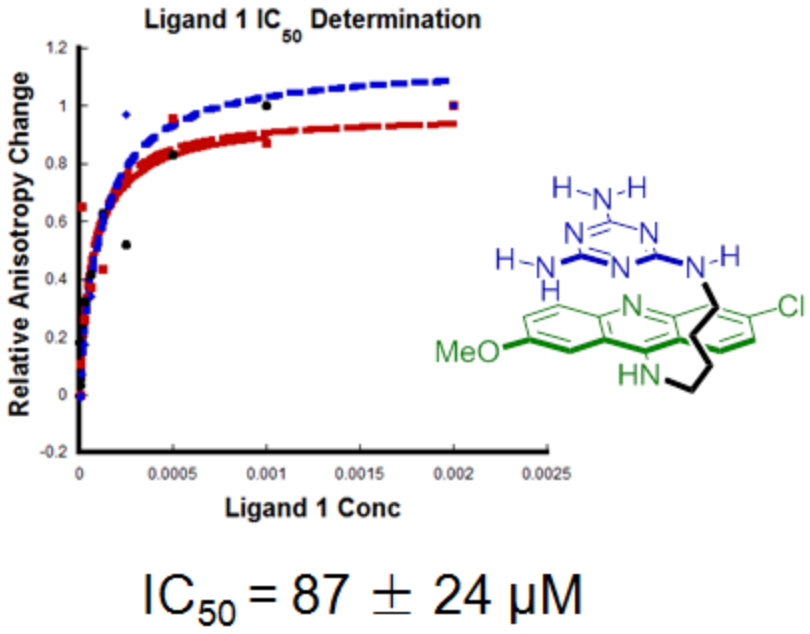


Figure 21. Fluoresence anisotropy data for ligand 1

Prior to re-screening the 117 compounds selected from the *in silico* screen, it was necessary to determine the appropriate conditions with the assay performed in a high-throughput format. Several factors needed to be addressed such as the order of addition and the incubation time for the small molecule with the RNA-protein complex.

Sample screening experiments were performed to determine the order of addition for the high-throughput screen, with the goal of minimizing addition steps. For small molecule inhibition studies there are two typical methods of addition; prevention mode where the RNA-small molecule is pre-incubated and protein is added later, and displacement mode where RNA-protein is pre-incubated with small molecules added to displace the complex. For gel shift experiments, prevention mode was employed because it resulted in better inhibitory response.

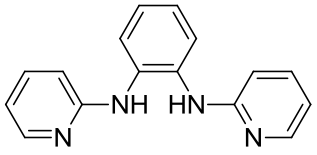
However, displacement mode requires one less compound addition step. In addition to reducing addition steps, displacement mode will reduce the potential for bubbling in the wells due to the presence of Triton-X. Triton-X is a detergent that is necessary to reduce small molecule aggregation but bubbles during any sample addition, leading to false fluorescence readings. From test screens we observed little difference in signal between the two methods of addition; thus, the displacement method was used.

The last factor addressed was the incubation time for the TAMRA-CUG_6, GST-MBNL1N and the small molecule. Time was an important factor for the folding of TAMRA-CUG_6 most likely due to a competition between duplex and stem-loop formation but it was not known whether assay response would change over time. Fluorescence anisotropy is an equilibrium assay so it is expected that response should not vary. We compared test screens measure at 20 minutes, 40 minutes and 1 hour and found that inhibitory responses did not change. One hour was selected as incubation time to reduce bubble formation.

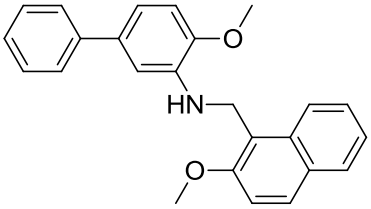
Identification of Lead Molecules from the Pharmacophore Screen

The screening of 117 compounds was carried out in a high-throughput manner in under 6 hours; including compound picking. Ligand **4** was used as a control and demonstrated a 47% signal change compared to DMSO control lanes. Overall, 18 compounds were identified as hits in the screen by demonstrating a greater than 10% signal change relative to the control (Figure 22). Of the 18 compounds identified, 3 compounds were previously identified from the EMSA CUG_{12} screen, labeled with (12), and 3 compounds were previously identified from the EMSA CUG_4 screen, labeled with (4). The binding affinity of these compounds has been determined by Stacie Richardson and John Craffey and unfortunately the inhibition of MBNL1 is quite weak,

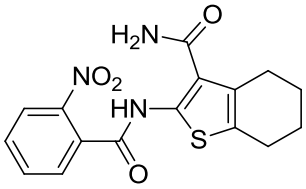
millimolar range. A future project would be the derivatization of such scaffolds to increase inhibitory activity.



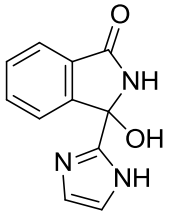
Chembridge 5367911 (12**)**



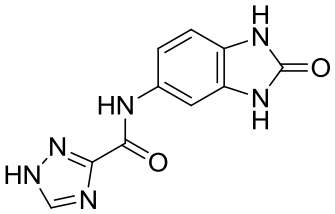
Chembridge 5457436



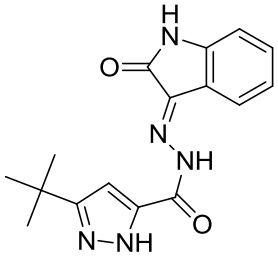
Chembridge 5571606



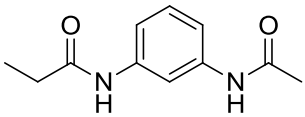
Chembridge 5735170



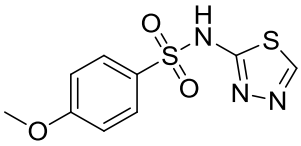
Chembridge 5738653



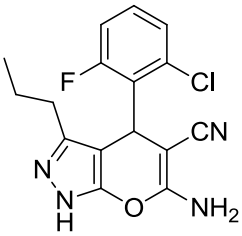
Chembridge 5784159



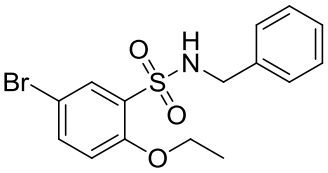
Chembridge 5836613



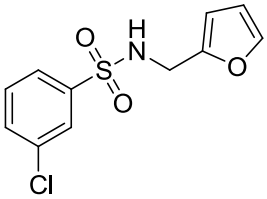
Chembridge 6595394



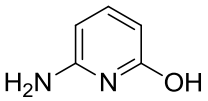
Chembridge 6659423 (12)



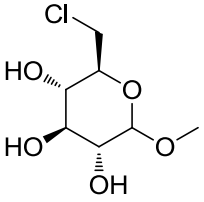
Chembridge 7240162



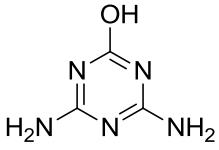
Chembridge 7434823



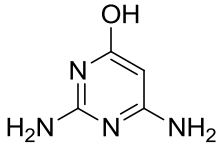
Marvel 531 (4)



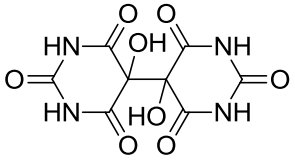
Marvel 1312



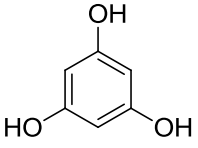
Marvel 3884



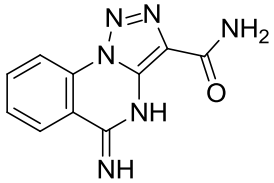
Marvel 4742 (4)



Marvel 4757 (12)



Marvel 4796 (4)



NSC 326654

Figure 22. Results from pharmacophore-based screen

High-throughput Screening of NCI Diversity Set III

Multiple lead compounds were identified from the *in silico* pharmacophore-based screen but most exhibited weak inhibitory activity. To identify more effective lead compounds, we decided to explore an expanded molecular space through a high-throughput screen of the NCI Diversity Set III library that contains over a million pharmacophores. This is possible because we have optimized the conditions of the fluorescence anisotropy assay for high-throughput screening. Steps in the screening process include an initial high-throughput fluorescence anisotropy screen of the whole library; a gel shift confirmatory screen; and a third screen in the presence of t-RNA. (Figure 23)

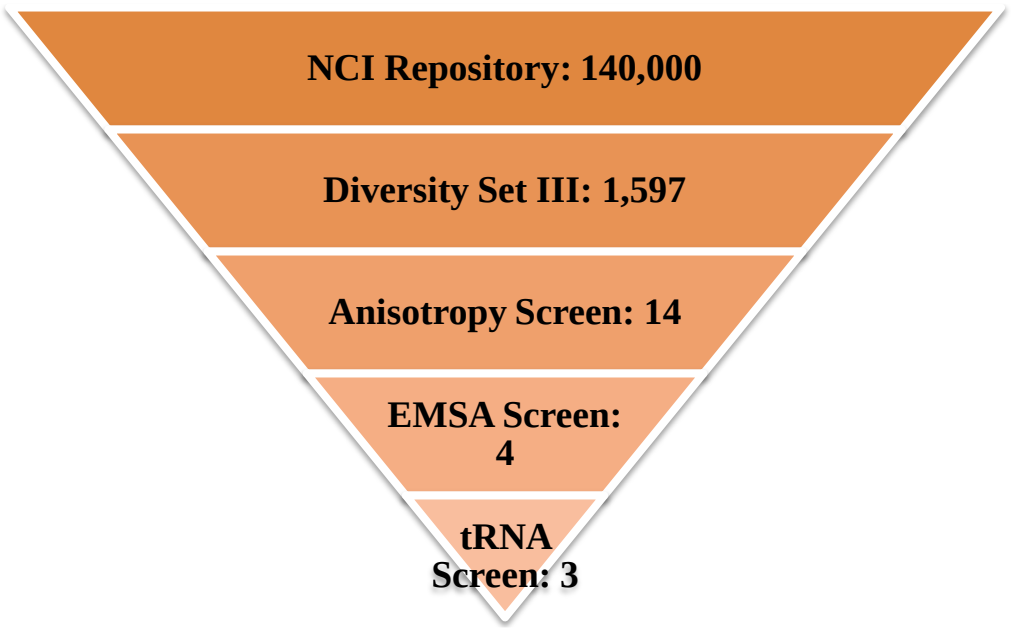


Figure 23. Summary of screening process with number of compounds at each level listed

Small Molecule Library Selection

The small molecule library Diversity Set III from the National Cancer Institute (NCI) was chosen for the study because it was easily accessible, inexpensive, and designed to represent a maximum amount of pharmacophores. From the 140,000 compounds stored at NCI, the 1,597 compounds of Diversity Set III were selected based on 1) compound availability, 2) representation of novel pharmacophore, 3) structural rigidity, 4) pharmacologically desirable features, and 5) purity of compound. The smaller size of the library is suitable for initial assay development. Upon identification of a hit compound, the NCI freely provides up to 10 mg of any compound listed in the Diversity Set for additional studies, expediting the discovery process. Compared to commercial small molecule libraries, the cost of using the Diversity Set is greatly reduced.

Screen Optimization and Results

Prior to screening the small molecule library the robustness of the assay was determined by calculating the Z-factor.[5] Z-factor is a mathematical description of the separation between the positive and negative controls ranging from 1 (ideal separation) to 0 (overlapping values for the control). A Z-factor of 0.5 to 1.0 is considered to be an excellent assay. In figure 7, the y axis is the anisotropy value for each of the tested compounds which are on the x axis. A total of 1,597 compounds were tested. In experiments with protein and RNA in buffer 0.178 ± 0.006 anisotropy value was obtained; representing a fully bound state. (Figure 24 red bar) In experiments with RNA only, representing a fully dissociated state, 0.121 ± 0.003 anisotropy value was obtained (Figure 24 green bar). From these values, a Z-factor of 0.53 was calculated which means that the assay is considered to be robust.

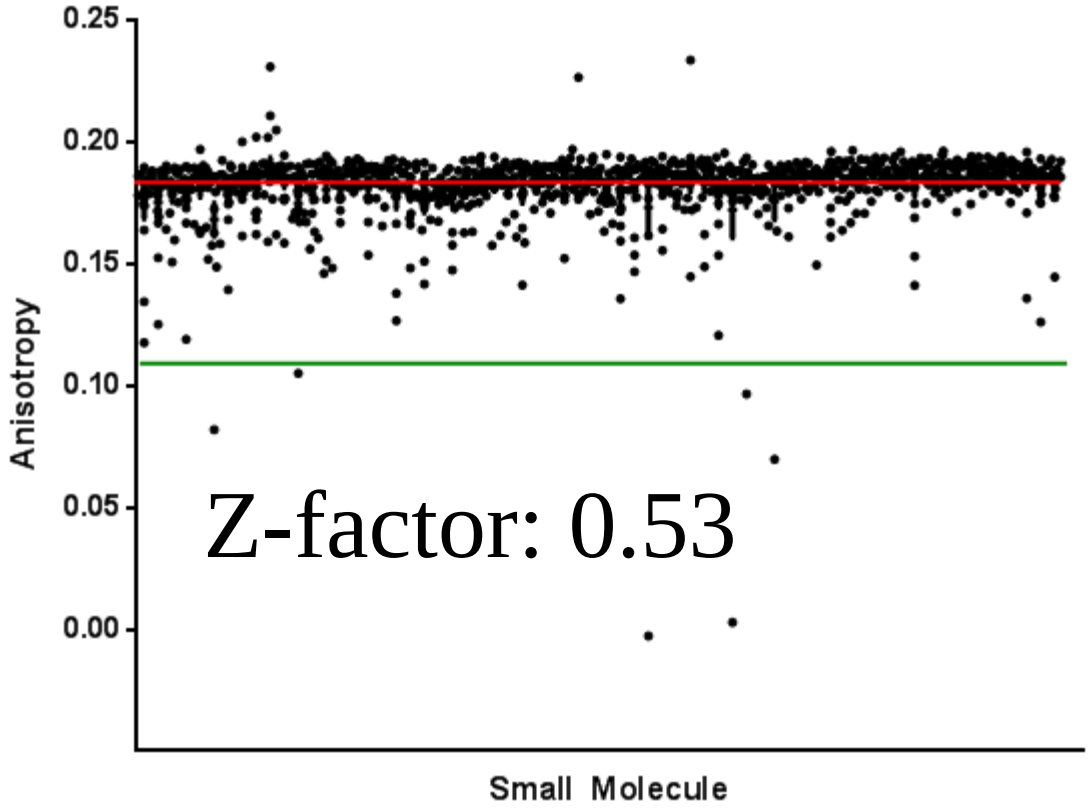
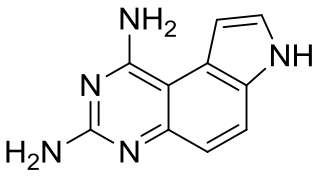


Figure 24. Plot of anisotropy data for all small molecules

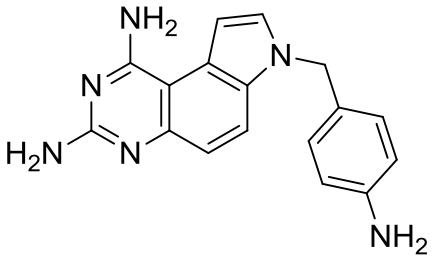
The high-throughput fluorescence anisotropy screen was performed at a small molecule concentration of 100 μM and resulted in the identification of 14 hits. Anisotropy is a measure of the loss of polarization of a fluorophore upon molecular rotation, thus larger complexes exhibit high anisotropy while smaller complexes exhibit low anisotropy. Potential inhibitors are expected to decrease the anisotropy value in a fluorescence anisotropy assay because it is preventing the formation of the larger complex. Compounds with anisotropy values below 0.14 were selected for follow-up studies to confirm the results. False positives are present in all high-throughput assays. These are shown as the data points that are outliers in figure 24. In the fluorescence anisotropy assay, false positives mainly result from 1) aggregation-based inhibition and 2) direct quenching of the fluorophore through small molecule-fluorophore interactions. To

address these issues the detergent Triton-X100 was used to reduce small molecule aggregation and a second confirmatory screen based on displacement of radio-labeled poly$(CUG)_{12}$ RNA instead of fluorophore-based detection was employed.

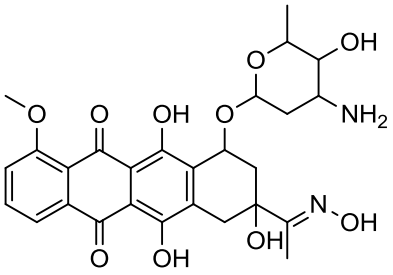
From the 14 hits previously identified, only 4 compounds remained active. (Figure 25) Ligand **13** and a series of derivatives were previously described as an inhibitor of dihydrofolate reductase in 1994.[6] Several derivatives, such as ligand **14** were reported to show anti-cancer activity in mouse models. Ligand **15** is a doxorubicin derivative, which is a known DNA intercalator.[7] Ligand **16**, also known as gentian violet, is a dye used in Gram staining. The same compound was identified as a hit molecule in a patent filed by Ohno and coworkers employing a high-throughput screen of biologically active compounds based on rescue of missplicing of cTnT in HEK293 cells.[8] Ligand **17** and a series of bis-quaternary salts were also investigated as an anti-tumor agents in 1974.[9]



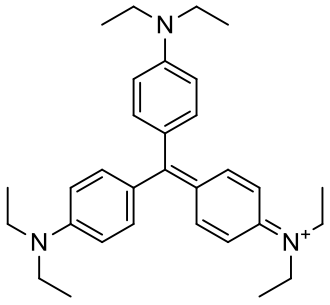
Ligand 13



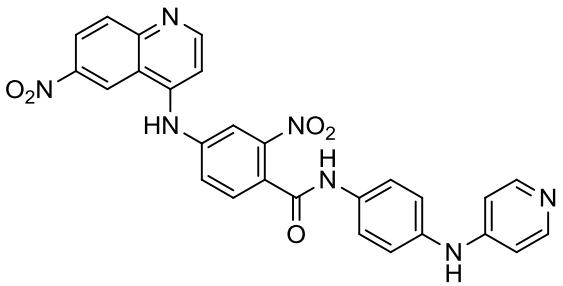
Ligand 14



Ligand 15



Ligand 16



Ligand 17

Figure 25. Ligands identified through screening and confirmed with gel shift

A third screen was employed to eliminate nonselective compounds by the addition of competitor tRNA. Increasing amounts of tRNA were included in gel shift assay at a ratio of 1500:1 and 5000:1 tRNA to poly$(CUG)_{12}$ RNA. Under such conditions, ligand **16** displayed loss of inhibition and was eliminated from the study. Of the remaining four lead compounds, the current focus is with ligand **13** and its derivative **14** because of a simpler synthetic scheme. Compounds **15** and **17** will be investigated in future work.

SAR with Similar Compounds

A structure-activity-relationship experiment was employed to understand the molecular motifs important for inhibition. Compounds similar to **13** defined by the Tanimoto similarity index were obtained from the NCI small molecule repository. The Tanimoto similarity index uses atom connectivity to determine compound similarity with values ranging from 1 being identical to 0 being completely dissimilar. Seventeen compounds were obtained from NCI with a similarity index higher than 0.9 but only one was a direct derivative of the compound **13**, compound **14**. (Figure 26) The goal of the experiment is to determine what structural motif of **13** is required for inhibition**.** Surprisingly, only **13**, ($P = 0.002$) and **14** ($P = 0.03$) demonstrated significant inhibition compared to the DMSO control whereas the other similar compounds did not exhibit significant activity (Figure 27). Compound **4** is the positive control and is one of the methylated derivatives of triaminotriazine-acridine conjugate. Interestingly, the hydrogen bonding motif of ligand **13** is similar to hydrogen bonding formed by triaminotriazine. (Figure 28) The presence of the donor acceptor donor acceptor (DADA) hydrogen bonding motif formed by the two amines on the pyrimidine present in **13** and **14** may be necessary for activity.

Compared to the parent compound, **14** demonstrated increased inhibition (P = 0.015) suggesting additional derivatization at the indole nitrogen may lead to improved inhibition activity.

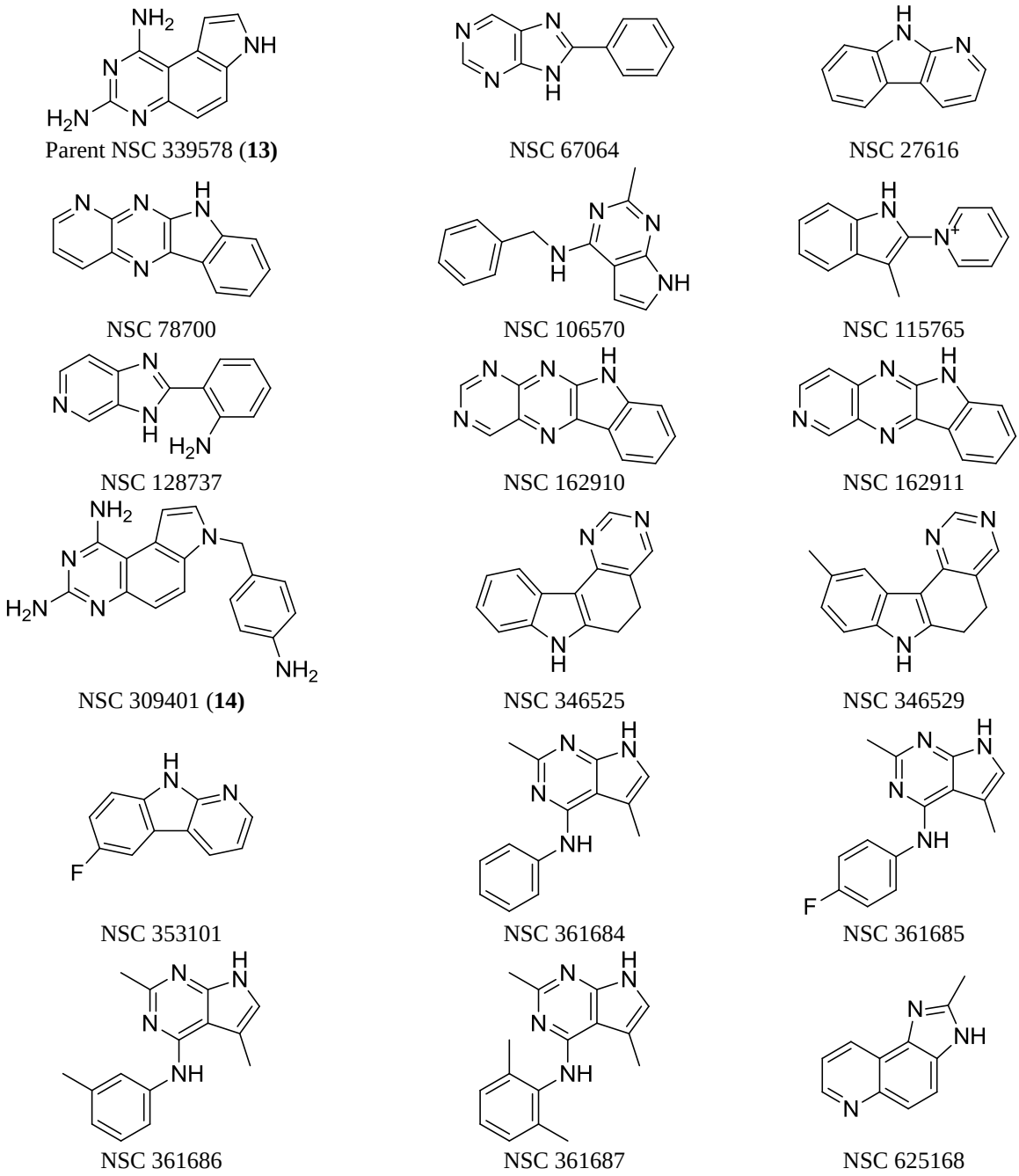


Figure 26. Compounds similar to ligand 13

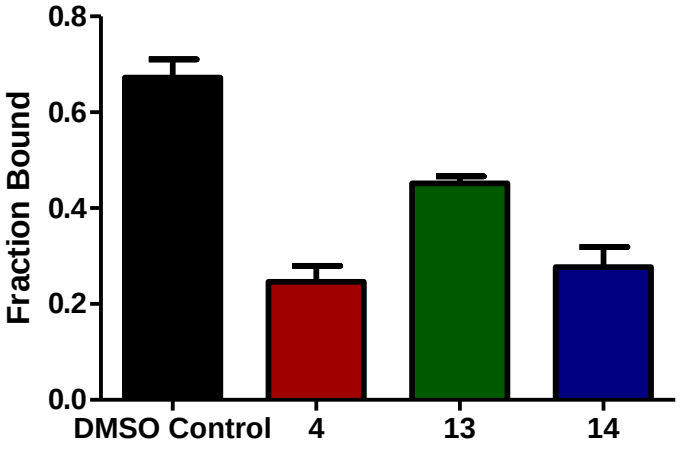


Figure 27. Summary of three independent SAR screening experiments (SEM is shown)

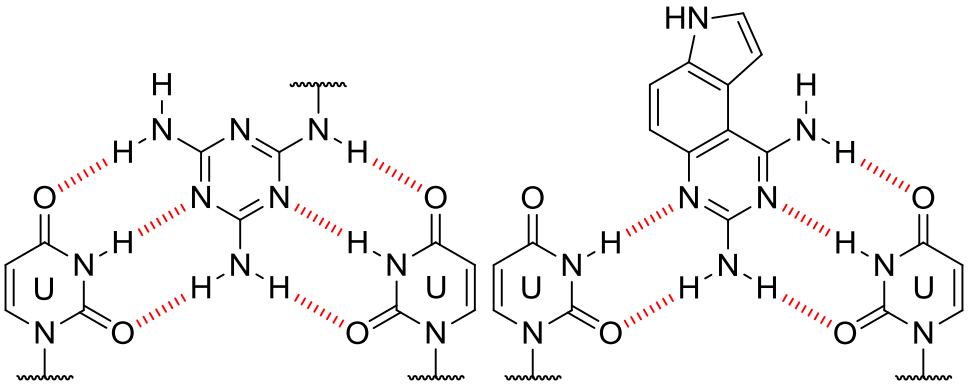


Figure 28. Comparison of h-bonding motif (Left) triaminotriazine (Right) compound 13

Dose Dependent Inhibition Activity of Ligand 13

Dose-dependence studies were completed with compound **13** via gel shift assay to establish a value for head-to-head comparisons against known small molecule inhibitors and future synthesized derivatives. Compound **13** exhibited an IC_{50} of 61 ± 15 μM with a calculated K_i of 15 ± 3 μM. (Figure 29) These values compare favorably to other small molecule inhibitors of DM1 such as pentamidine with an IC_{50} of 58 ± 5 μM and compound **4** with an IC_{50} 48 ± 1 μM.[4,10] Dose dependence studies with **14** were inconclusive because of small molecule aggregation at high small molecule concentrations leading to an incomplete dose-dependence profile. The solubility of **14** will need to be characterized prior to additional dose-dependence experiments. The low micromolar IC_{50} suggests **13** is a promising scaffold for further derivatization but compound solubility must be considered in derivative design.

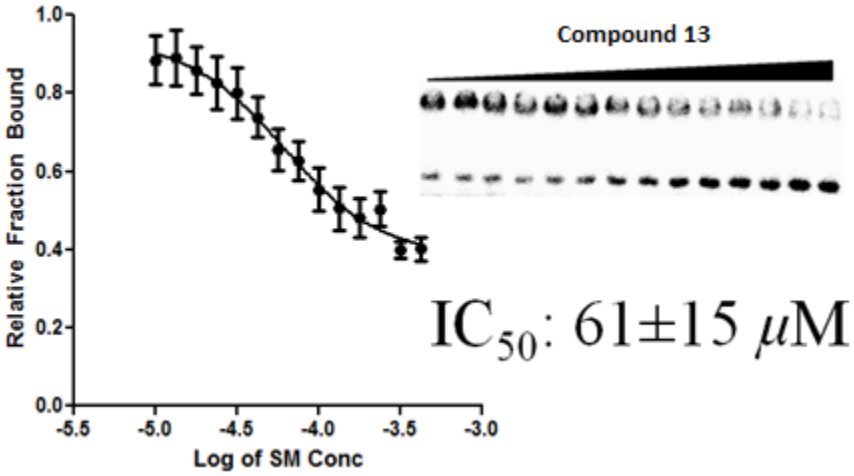
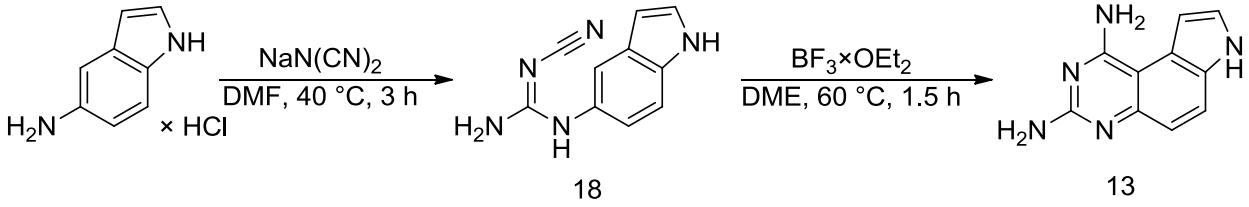


Figure 29. Representative gel and densitometry data for 7 independent dose-dependence experiments

Synthesis of 13

The synthesis of 2,4-diaminopyrroloquinazoline, **13**, was first described by Ledig and coworkers through a thermally-induced condensation of sodium dicyanamide and 5-aminoindole.[11] In 1994 Styles and coworkers reported an improved synthesis through the isolation of a cyano intermediate **7** followed by $BF_3{\cdot}OEt_2$ assisted cyclization. (Scheme 1)[12] Following this method intermediate **18** has been successfully synthesized in high yield. Synthesis of compound **13** and its derivatives is now the project of a new graduate student.



Scheme 1. Synthesis of ligand 13

3.3 Conclusions

The fluorescence anisotropy assay was developed to provide another method to analyze the lead molecules identified from the *in silico* pharmacophore-based screen. However, a general screen of the NCI Diversity Set III has been more successful in identifying strong inhibitors of poly$(CUG)_n$-MBNL1 complex than the pharmacophore-based screen. We can now take advantage of the high-throughput capabilities of the anisotropy assay to quickly identify multiple lead compounds and to expand our portfolio of small molecule inhibitors for myotonic dystrophy.

In regards to the identified molecules, additional experimental data is required to demonstrate that the RNA is the target of the small molecule as opposed to the protein. Quantification methods such as ITC and RNA melting experiments will be necessary for demonstrating small molecule-RNA binding. Additional biological assays, such as fluorescence microscopy to visualize the displacement of mislocalized MBNL, splicing assays to determine the effects on splicing, and *in vivo* toxicology studies will be critical in evaluating the potential of the lead compound to be developed further.

The identification of active inhibiters possessing activities comparable to published small molecule inhibitors has shown the robustness of the fluorescence anisotropy assay. However, the success of all high-throughput screening experiments is limited by the choice of library. The larger the library, the more chemical space explored with the expectation of discovering more effective inhibitors. As a proof of concept experiment, the screening of Diversity Set III, a relatively small library, demonstrates the effectiveness of the method. Thus, future screens can employ much larger small molecule libraries such as the one available at the Molecular Libraries Program at NIH with 300,000 compounds with the goal of identifying even more potent lead compounds.

3.4 Materials and Methods

GST-MBNL1N Expression and Purification

Bl21 E. coli cells were transformed with the GST-MBNL1N plasmid. A mixture of 50 µl of cell, 1 µL of plasmid, and 0.7 µL of 10% BME was kept on ice for 20 min and then heat shocked for 45–55 sec at 42 °C. After returning to ice for 30 sec, 450 µL of SOC media was added at room temperature. The mixture was incubated in shaker at 37 °C for 1 h. and then plated on agar and placed in a convection oven at 37 °C overnight for bacterial growth.

Transformed bacteria were induced with IPTG, once the OD_{600} reached 0.6, for 2 h at 37 °C. The solution was centrifuged for 15 min at 10,000 g and the pellet was saved and treated with lysis buffer (tris-Cl pH 8, 10 mM imidazole, 2mg/mL lysozyme, 0.1% Triton X, 0.1% protease inhibitor, 0.5 M NaCl, 2 mM BME, 5%glycerol). Additionally, DNAse 1 was added to the lysate in a ratio of 1 µL of DNAse to 1 mL of lysis buffer. After 6 rounds of sonication at 10 sec each followed by 30 sec ice bath, the sample was centrifuged at 12,000 g for 15 min. The supernatant was filtered with a 0.45 µM Millex Filter (Millipore). Two mL of Glutathione Sepharose 4B (GE Healthcare) was incubated with the lysate for 1 h at 4 °C. The solution was loaded on a column and the beads were washed with 1) 25 mL of wash buffer A (25 mM tris-Cl pH 8, 5 mM BME, 1 M NaCl, 0.1% Triton X), 2) 15 mL of wash buffer B (25 mM tris-Cl pH 8, 5 mM BME, 300 mM NaCl, 0.1% Triton X), 3) 10 mL of wash buffer C (25 mM tris-Cl pH 8, 5 mM BME, 150 mM NaCl, 0.1% Triton X), and eluted with 20 of Elution Buffer 2 (50 mL of wash buffer C, 10 mM reduced glutathione). Two mL of Ni NTA Sepharose (QIAGEN) was added and incubated with lysate for 1 h at 4 °C. The solution was loaded on a column and the beads were washed with 20 mL of Wash Buffer1 (25 mM tris-Cl pH 8, 20 mM imidazole, 0.1%

Triton-X, 0.5 M NaCl, 5 mM BME) and eluted with 20 mL of Elution Buffer 1 (25 mM tris-Cl pH 8, 250 mM imidazole, 0.1% Triton-X, 0.5 M NaCl, 5 mM BME). The protein sample is dialyzed into storage buffer (100 mM NaCl, 25 mM tris-Cl pH 7.5, 5 mM BME, 25% glycerol) using molecular porous membrane tubing with molecular weight cutoff of 3.5 kDa (Spectra/Por) first for 1 hour and then overnight.

The concentration of protein obtained was determined with a BCA assay (Thermo Scientific) while purity was confirmed with MALDI mass spectrometry and SDS-PAGE.

Fluorescence Anisotropy Equilibrium Binding Assay

The fluorophore is protected from light with aluminum foil during all steps. TAMRA-CUG_6 was diluted with half the necessary amount of buffer with Anisotropy Folding Buffer (20 mM tris-Cl pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM BME, 0.05% Triton-X). Next the sample is heat shocked at 90° C for 5 min and then put on ice for 30 min. Anisotropy Buffer (20 mM tris-Cl pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM BME, 0.05% Triton-X, 2 mg/mL BSA) is added to reach a final concentration of 40 nM.

The protein was serially diluted with anisotropy buffer. The two were mixed and incubated at room temperature for 20 min before being transferred onto a 384-well plate (Greiner). Fluorescence anisotropy readings were measured on an Analyst HT instrument using a beam splitter at 561 nm with excitation at 530±12.5 nm and emission at 570±5 nm. Apparent K_d was obtained by graphing the change in anisotropy and fitting to the equation anisotropy = $1/(1 + K_d / [\text{Protein}])$ with KaleidaGraph 3.5.

Fluorescence Anisotropy Inhibition Screening Assay

The RNA was prepared in the same fashion as the fluorescence anisotropy equilibrium binding assay. The protein concentration used is the 80% bound state relative to the K_d. Loading volumes were 10 μL with 4.5 μL of RNA, 4.5 μL of protein, and 1 μL of small molecule. The small molecule was first loaded onto the 384-well plate (Greiner) and a 20 min pre-incubated mixture of TAMRA-CUG_6 and GST-MBNL1N was added. After an incubation of 1 h, fluorescence anisotropy readings were measured on an Analyst HT instrument using a beam splitter at 561 nm with excitation at 530±12.5 nm and emission at 570±5 nm. Inhibition was determined by comparing anisotropy change relative to DMSO control.

Fluorescence Anisotropy IC_{50} Assay

The RNA was prepared in the same fashion as the fluorescence anisotropy equilibrium binding assay. Small molecules were serially diluted from 10 mM and loaded onto the 384-well plate (Greiner). TAMRA-CUG_6 and GST-MBNL1N were incubated for 20 min prior to addition onto plate. After an additional 1 h incubation fluorescence anisotropy readings were measured on an Analyst HT instrument using a beam splitter at 561 nm with excitation at 530±12.5 nm and emission at 570±5 nm. To determine IC_{50} values the data was fit to the equation $B = \Delta B \exp((-0.69/IC_{50})C) + B_f$ where B is the relative anisotropy value and ΔB is the difference in the anisotropy value at the lowest and highest concentrations of small molecule, C is the concentration of the small molecule, and IC_{50} is the concentration of small molecule at which the relative anisotropy value is half the difference of the lowest and highest values.

High-throughput Anisotropy Screen

TAMRA-CUG_6 was diluted with half the necessary amount of buffer with Anisotropy Folding Buffer (20 mM tris-Cl pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM BME, 0.005% Triton-X). Next the sample is heat shocked at 90^0 C for 5 min followed by slow cooling over 30 min to RT. Anisotropy Buffer (20 mM Tris-Cl pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM BME, 0.005% Triton-X, 2 mg/mL BSA) is added to reach a final concentration of 44.44 nM. GST-MBNL1N was diluted to 1.11 μM with anisotropy buffer. The two were mixed and incubated at room temperature for 20 min before being transferred onto 5 384-well plates (Greiner) using Matrix Platemate (Thermo Sci.). Each well was contained 9 μL of pre-incubated RNA and protein. Pin transfer was used for the addition of 1 μL of 1 mM Diversity Set III stock to each well. Fluorescence anisotropy readings were measured on an Analyst HT instrument using a beam splitter at 561 nm with excitation at 530±12.5 nm and emission at 570±5 nm.

3.5 References

1. Mao, C; Flavin, KG; Wang, S; Dodson, R; Ross, J; Shapiro, DJ. Analysis of RNA-protein interactions by a microplate-based fluorescence anisotropy assay, *Anal. Biochem.* **2006**, *350*, 222–232.

2. Thornton, C. *Pubchem Bioassays*. "qHTS Assay for Inhibitors of MBNL1-poly(CUG) RNA binding: Initial hit validation in AlphaScreen Assay." Accessed August 2, 2011

3. Einarson, MB; Pugacheva, EN; Orlinick, JR. Preparation of GST Fusion Proteins, *Cold Spring Harb Protoc*, **2007**. doi: 10.1101/pdb.prot4738

4. Zimmerman, SC; Arambula, JF; Ramisetty, SR; Baranger, AM. A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 16068–16073.

5. Zhang, J H; Chung, TD; Oldenburg, KR. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays, *J Biomol Screen*. **1999**, *2*, 67–73.

6. Stables, JN; Kuyper, LF; Baccanari, DP; Jones, ML; Hunter, RN; Tansik, RL; Joyner, SS; Boytos, CM; Rudolph, SK; Knick, V; Wilson, HR; Caddell, JM; Friedman, HS; Comley, JC. High-affinity inhibitors of dihydrofolate reductase: antimicrobial and anticancer activities of 7,8-dialkyl-1,3-diaminopyrrolo[3,2-f]quinazolines with small molecular size, *J. Med. Chem.* **1996**, *39*, 892–903.

7. Di Marco, A; Zunino, F; Silvestrini, R; Gambarucci, C; Gambetta, RA. Interaction of some daunomycin derivatives with deoxyribonucleic acid and their biological activity, *Biochem. Pharmacol.* **1971**, *20*, 1323–1328.

8. Ohno, K. Matsuura, T. National University Corporation Nagoya University. Agent for Treatment of Myotonic Dystrophy. International Patent. 062054. 01.20.2011

9. Cain, BF; Atwell, GJ. Potential antitumor agents. 15. Bisquaternary salts, *J. Med. Chem.* **1974**, *17*, 930–934.

10. Warf, B; Nakamori, M; Matthys, CM; Thornton, CA; Berglund, JA. Pentamidine reverses the splicing defects associated with myotonic dystrophy, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 18551–18556.

11. Ledig, KW. 7-(Substituted)-7H-pyrrolo[3,2-f]quinazoline-1,3-diamines. U.S. Patent 4,118,561, 1978

12. Jones, ML; Kuyper, LF; Caddell, JM; Styles, VL. Lewis acid assisted cyclization of arylcyanoguanidines to 2,4-diaminoquinazolines, *J. Heterocyclic Chem.* **1994**, *31*, 1681–1683.